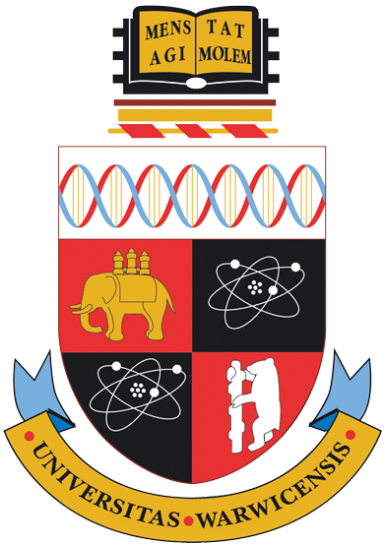

Studies of rare $B \to K\pi\mu\mu$ decays using the LHCb experiment

by

Andrew Crocombe

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

Department of Physics

University of Warwick

September 2018

Contents

Acknowledgments

First and foremost, I would like to thank Tom Blake for his excellent advice, constant support and his seemingly endless patience. I could not have hoped for a better supervisor over the past four years. I would also like to thank Tim Gershon and Michal Kreps for my initial introduction to the world of flavour physics and their continued support when needed.

I thank Mark Whitehead for the help, career advice and sharing of beer and crispy patas over the years. I also express my admiration for his ability to put up with me and Dan during our lunchtimes in Geneva. I extend a similar thank you to Tom Latham for all his help and my commiserations for the times spent lost in the VELO geometry description with me and Abhijit.

To everyone else I've had the pleasure of knowing in the Warwick LHCb group, my thanks for making sure the work environment was always interesting. In particular, I have thoroughly enjoyed our shared trivia experiences and my apologies for being one of the fools who broke the system.

This thesis would not have been possible without the support of so many of my friends, both inside and outside of physics. I thank them for sharing in a variety of quizzes, the many games of pool, the visits to the real ale festival, the frankly ridiculous number of bottles of wine shared, the opportunity to make homemade gin and mozzarella and for always believing that I could do it.

Finally, I would like to thank my brother, parents and grandparents for always showing an interest in what I've been doing, their unconditional support through every challenge and for all the words of encouragement that have helped along the way.

Declaration

I declare that the work presented in this thesis is my own original work except where explicitly referenced. The details of any previous publications resulting from this work are giving within the body of the text. This thesis has not been submitted, in any form, for another degree from this or another university.

Abstract

Within the Standard Model of particle physics, decays involving flavour-changing neutral current transitions cannot occur via tree-level Feynman diagrams and can only occur at loop level or higher. This leads to a heavy suppression of decay modes of this type and makes them sensitive to contributions from particles beyond the Standard Model that can enter in competing Feynman diagrams. By performing measurements of the rate and angular distribution of these decay modes, powerful tests of the Standard Model can be carried out.

In this thesis, a first search for the decay $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ is presented using data samples collected by the LHCb collaboration during Run 1 and Run 2 of the LHC. An excess of this decay is found over the background-only hypothesis, providing first evidence for the decay mode with a significance of 3.4 standard deviations. A first measurement of the branching fraction of this decay is made and a discussion of the future prospects for measurements of the decay mode is presented.

Chapter 1

Introduction

This thesis describes the main contributions I have made to the LHCb experiment over the course of my Ph.D. The main focus of this work is an analysis to search for the rare decay $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$, where the inclusion of the charge conjugate process of this and other decay modes is implied throughout this thesis unless otherwise specified. The search for the decay mode $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ detailed in this thesis has previously been published in Ref. [1]. In addition to this analysis, I have worked on studies that have helped to make design choices for the upgrade of the LHCb experiment, to be carried out after Run 2 of the Large Hadron Collider (LHC). An overview of these studies is given in the appendix of this thesis. The work described here is not a complete record of my contribution to the LHCb experiment as a member of the collaboration. I have also participated in the data collection effort of the experiment during Run 2 of the Large Hadron Collider (LHC). I have mainly contributed as an on-call expert for the vertex locator (VELO) sub-detector and as a developer of software designed to automatically monitor the data quality of the VELO over time.

The structure of this thesis, with the main focus being on the analysis of the decay mode $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$, follows. In Chapter 2, an overview of the Standard Model (SM) of particle physics is given, with particular focus on how observables in rare heavy flavour physics can be used to search for evidence of new particles arising in extensions of the SM. This is followed by a brief experimental overview of the current status in rare heavy flavour physics. A description of the Large Hadron Collider and its suitability as an environment to make precise experimental measurements of heavy flavour physics decays is given alongside a detailed description of the components and operations of the LHCb detector in Chapter 3. An introduction to the rare flavour-changing neutral current decay $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ along

with the analysis strategy for searching for this decay with the LHCb detector is given in Chapter 4. This is followed by details on the candidate selection and an overview of backgrounds considered in the analysis in Chapter 5. Chapter 6 describes the calculation of efficiencies of decay modes needed to make a branching fraction measurement and Chapter 7 provides the details of an invariant mass fit used to extract the yield of $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ decays as well as the calculation of its branching fraction. Full details of the systematic uncertainties on the branching fraction are given in Chapter 8 and a summary of the results of the analysis and a discussion on future prospects of other measurements of the decay $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ are given in Chapter 9. Finally, Appendix A provides the details of studies carried out to assist in design choices for the upgraded LHCb detector VELO.

Chapter 2

Theoretical overview

In this chapter, the theoretical framework of the Standard Model of particle physics is described and detailed overviews of the key components related to heavy flavour physics are given. Rare decays are described as tools for searching for physics beyond the SM and the framework in which this new physics (NP) can be parameterised is detailed alongside details of observables sensitive to NP. Finally, an overview of the current experimental status of rare decay measurements at the B factories and the Large Hadron Collider is given. Where not otherwise referenced, details in this chapter are taken from Refs. [2, 3].

2.1 The Standard Model of particle physics

All of the known particles in the universe and their dynamics can be described by the SM [4–7], a quantum field theory (QFT) that describes interactions via the electromagnetic, strong and weak nuclear forces. The SM is one of the most successful models in physics, being both predictive and robust under all direct measurements carried out to date.

In Fig. 2.1 all of the known particles in the SM and their fundamental properties are shown. These particles can be split into two categories depending on whether they have integer or half-integer spin, called bosons and fermions respectively. The vector bosons, with spin-1, are the force mediating particles in the SM, with the photon (γ) mediating the electromagnetic force, the $W^{\pm}$ and Z^0 mediating the weak force and the gluon (g) mediating the strong force. The final boson is the Higgs boson, a spin-0 particle that arises as a part of the Higgs mechanism through which mass is generated for the massive particles.

Fermions are the particles which make up the matter in the universe and can

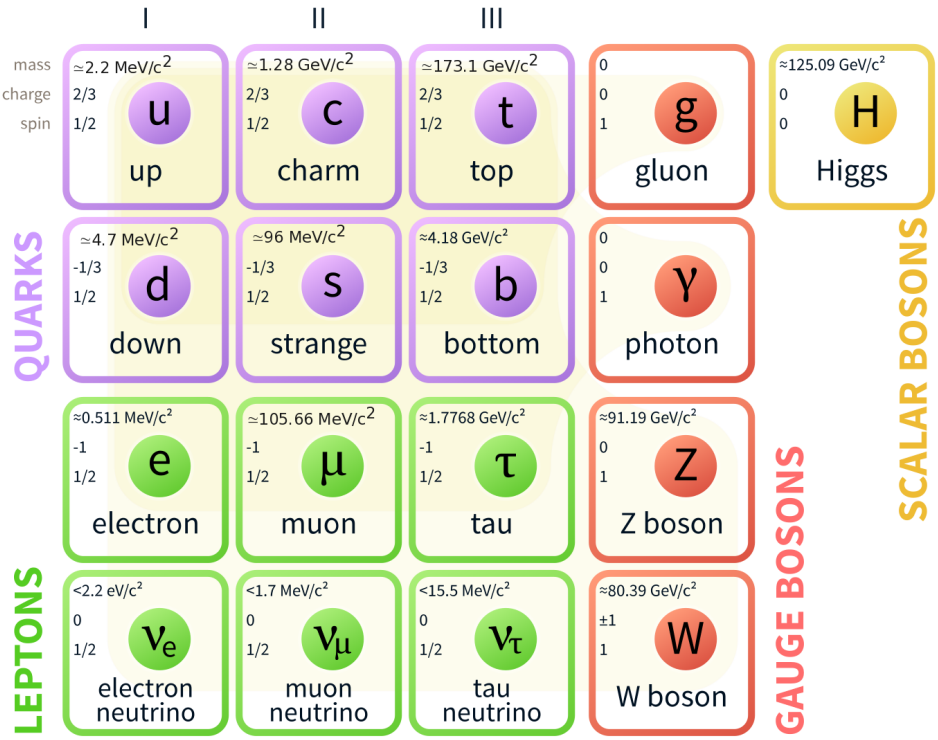

Figure 2.1: Schematic representation of the Standard Model of particle physics, detailing the constituent and force mediating particles and well as their masses, charges and spins. Reproduced from Ref. [8].

be further divided into quarks and leptons. The most common of these particles are the up-quark, down-quark, electron and electron neutrino, with these four making up the vast majority of matter in the universe. The remaining fermions are essentially higher mass versions of these most common particles, organised into two further generations with the same structure as the first. Generation two is made up of the charm-quark, strange-quark, muon and muon neutrino and generation three the top-quark, bottom-quark, tau and tau neutrino. In addition to these three generations of particles, there exists the anti-matter version for each of these particles with opposite charge and parity.

The quarks are the only particles to interact via the strong nuclear force and have an additional property known as "colour", with three possible states: red, blue and green as well as the corresponding anti-versions for the anti-quarks. Quarks are not found in a singular state and are required to form colourless bound states, either requiring one quark (or anti-quark) with each of the three colours (or anti-colours) or a pair of a quark and anti-quark with the matching colour and anti-colour. These three quark states are referred to as baryons while the quark and anti-quark pairs

are referred to as mesons. It is also possible to make colourless combinations of quarks and anti-quarks with more than three constituents, with four [9–13] and five [14] constituent states having been observed.

The force mediating bosons in the SM are associated with the fundamental interactions. All of these interactions are dictated by symmetry principles and they are associated with the gauge symmetry group described by

$$SU(3)_C \times SU(2)_W \times U(1)_Y, \tag{2.1}$$

where C denotes colour, W weak isospin and Y hypercharge. One of the key principles of the of SM is the requirement of local gauge invariance of the generator symmetry groups. This is an invariance under a local transformation of the fields. The different parts of this gauge symmetry group describe the different fundamental interactions in the SM. The $SU(3)_C$ symmetry is associated with quantum chromodynamics (QCD) and the strong interaction via gluons. The combination of the weak and electromagnetic forces, the electroweak (EW) force, is associated to $SU(2)_W \times U(1)_Y$.

To describe the interactions of the SM fields and particles, the total Lagrangian

$$\mathcal{L}_{\text{SM}} = \mathcal{L}_{\text{EW}} + \mathcal{L}_{\text{QCD}} + \mathcal{L}_{\text{Higgs}}, \tag{2.2}$$

is used, where the dynamics of the individual components of the SM can be detailed separately.

2.1.1 The Higgs mechanism

The requirement of gauge invariance means that the inclusion of masses in the electroweak sector is not simple. While the photon is massless, the other mediating particles are not and a mechanism for including mass terms is required. This is provided by the concept of electroweak symmetry breaking (EWSB) [15–18]. This is done with the addition of the scalar Higgs field, ϕ, with a non-zero vacuum expectation value (VEV).

The Lagrangian for this additional field is given by

$$\mathcal{L}_{\text{Higgs}} = (D_\mu \Phi)^\dagger (D^\mu \Phi) - V(\Phi^\dagger \Phi), \tag{2.3}$$

where Φ is a doublet of complex scalar fields that can act on the $SU(2)$ group

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}. \tag{2.4}$$

The Higgs potential from this Lagrangian is given by

$$V(\Phi^\dagger \Phi) = \mu^2 \Phi^\dagger \Phi + \frac{1}{2}\lambda(\Phi^\dagger \Phi)^2 \tag{2.5}$$

where μ is related to the mass of the Higgs boson and λ is the Higgs self coupling. When $\mu^2 < 0$, this Higgs potential has a non-zero minima at $v^2 = \mu^2/\lambda$, the VEV, which lies in a band in the complex plane around zero. At this point any particular minima from this band can be chosen. By taking the minima that only has a component for the real, neutral part of the doublet Φ, *i.e.* $\phi_3^2 = v^2$ with the other ϕ equal to zero, the value of Φ at the minima is

$$\Phi_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}. \tag{2.6}$$

By expanding the field around this particular non-zero minima, the symmetry of the field can be spontaneously broken while retaining gauge invariance. After substituting the chosen value in the Lagrangian, terms with the same form as the mass term of a boson are found. This breaking of the symmetry leads to mass terms for the weak mediating particles, $W^\pm$ and Z^0, while keeping the photon without a mass term (as the vacuum is still invariant under the subset of transformations related to electromagnetism, $U(1)_{EM}$).

The substitution of the VEV into the Higgs Lagrangian generates direct mass terms for the charged vector bosons, $W^\pm$, as well as diagonal matrix terms for the neutral vector bosons. By diagonalising the coupling of the Higgs field to the $SU(2)_W \times U(1)_Y$ gauge fields, physical fields corresponding to the neutral vector bosons are generated from the mixing, with corresponding mass terms. One of the eigenvalues of the matrix these terms is taken from is zero and this leads to the massless eigenstate that is the photon, A_μ, as well as the massive Z_μ.

As a consequence of the procedure that generates the mass terms for the mediating bosons, the remaining Higgs field is a real scalar field with a proper mass term and self-interactions. As such, the presence of a massive Higgs boson has been one of the most important testable predictions of the SM. The discovery of the Higgs boson by ATLAS [19] and CMS [20] was an important verification of the mechanism

of EWSB.

2.1.2 The Yukawa mechanism

The Higgs mechanism has provided an avenue through which the masses of the bosons can be generated, it also can do the same for the fermions. There are no direct mass terms for the fermions, however there are allowed Yukawa interaction terms between the Higgs doublet and quark doublets (in pairs of the same generation) and singlets. The Lagrangian for this is given by

$$\mathcal{L}_Y = -Y^d_{ij}\overline{Q^I_{Li}}\phi d^I_{Rj} - Y^u_{ij}\overline{Q^I_{Li}}\epsilon\phi^* u^I_{Rj} + \text{h.c.}, \tag{2.7}$$

where Y are the 3×3 complex Yukawa coupling matrices, Q_L are left-handed quark doublets, d_R and u_R are right-handed quark singlets and ϵ is the 2×2 antisymmetric tensor.

With the Higgs field acquiring a VEV as described above, these Yukawa couplings combined with the inserted VEV become mass terms for each of the quarks, $m_{u,d} = Y^{u,d}(v/\sqrt{2})$. To obtain the physical mass states for the quarks, these mass terms need to be diagonalised via unitary rotation matrices. By carrying this procedure out, the coupling between the physical quark states and the charged-current weak interaction becomes parameterised by a single matrix describing flavour mixing between quark generations.

2.1.3 The CKM matrix

The weak charged-current interaction is the only force which allows the changing of quark flavour in the SM. The couplings between the quarks and the weak force are quantities in the SM that are not predicted and are free parameters. All of these couplings can be described by the single 3x3 unitary matrix described above, parametrised by 3 mixing angles and a single complex phase, the CKM matrix [21, 22]. A general representation of this matrix is given by:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} = V_{CKM} \begin{pmatrix} d \\ s \\ b \end{pmatrix}. \tag{2.8}$$

As experimental measurements of the components have been made, it has become apparent that there is a hierarchal structure to the CKM matrix. One convenient parametrisation is to expand each element of the matrix as a power

series of V_{us}, this is the Wolfenstein parametrisation [23] and is given by:

$$V_{CKM} = \begin{pmatrix} 1 - \frac{\lambda^2}{2} & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \frac{\lambda^2}{2} & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4), \tag{2.9}$$

where $A \approx 0.8$ and $\lambda \approx 0.22$.

The presence of a complex phase in the CKM matrix has the important consequence that there is a difference between the flavour transitions involving quarks and anti-quarks. The difference between particles and anti-particles in the SM is described by the operation that transforms between the two, CP. This is the combination of the charge conjugation operation, C, which inverts the charge of a particle and the parity inversion operation, P. Both of the symmetries resulting from these operations are individually maximally violated by the weak interaction however the combination is generally a good symmetry. The complex phase allows for a small degree of CP violation within the SM, the amount of which cannot be derived from first principles but can be measured from CKM parameters. CP violation is one of the three Sakharov conditions necessary to generate a matter-antimatter asymmetry in the universe [24].

The unitarity of the CKM matrix can be represented as a triangle in the complex plane with an area that is a measure of SM CP violation. These are known as unitarity triangles with the most commonly used one being generated from

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0 \tag{2.10}$$

and where the terms are all divided by $V_{cd}V_{cb}^*$ to produce one side with length equal to unity.

As the CKM elements are fundamental parameters of the SM that cannot be derived, it is important to make the best possible measurements of them. This is also important to determine the amount of CP violation allowed within the SM. The angles of the unitarity triangle detailed above are essentially measurements of this CP violation. Over-constraining the measurements related to this triangle allow for powerful tests of the SM by testing the consistency between the measurements, where any inconsistency could be indicative of contributions from a NP process. The results of a global fit that takes into account a combination of the current best measurements and constraints related to the CKM matrix elements is shown in Fig. 2.2 as related to the unitarity triangle.

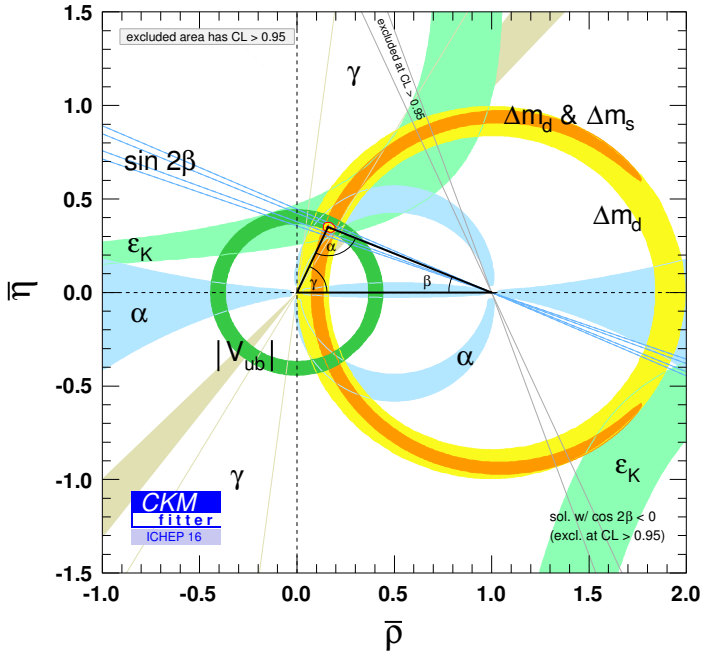

Figure 2.2: Constraints on the CKM parameters shown in the complex plane [25]. The matrix element $|V_{ub}|$ is extracted from decays of the type $B \to X\ell\nu_\ell$, ϵ_K from decays of neutral kaons, Δm_d and Δm_s from B^0 and B_s^0 oscillation measurements. The angle measurements of α are determined from the decays $B \to \pi\pi$, $\rho\pi$ and $\rho\rho$, β from decays of $B \to (c\bar{c})K$ and γ from tree level decays of $B^- \to D^{(*)0}K^{(*)-}$ versus $B^- \to \overline{D}^{(*)0}K^{(*)-}$.

2.1.4 Beyond the SM

While the SM is a highly successful model, being remarkably predictive while describing the fundamental interactions in physics, it is not a complete model of the universe. A number of key issues have been noted where there is no current explanation within the SM framework, which are detailed in the following.

Perhaps the largest deficiency in the SM is its inability to describe the large scale structure of the universe. Through a number of experimental observations, such as the rotation curves of galaxies [26] and the observation of gravitational lensing of astronomical objects [27], it is clear that the amount of visible matter in the universe is not large enough to account for the size of gravitational effects. From these measurements it would seem that the majority of matter in the universe is not visible, *i.e.* it does not interact with the electromagnetic force. This unaccounted for matter is known as dark matter, and it is unknown what it is composed of.

In addition, while the SM describes three of the fundamental forces to a great

precision, the fourth fundamental force, gravity, is not included. For a complete model of the universe the inclusion of all the fundamental forces in a quantum field theory is desirable. The scale at which quantum gravitational effects are expected to become important is known as the Planck scale and the masses involved are around 16 orders of magnitude greater then those generally involved with SM particles. The lack of a quantum gravitational theory is a shortcoming and this huge unexplored mass region presents plenty of opportunity for the presence of beyond SM particles.

Another puzzle is the observed matter-antimatter asymmetry in the universe. By measuring the number of observed baryons in the universe with respect to the number of photons, under the assumption they have been produced from matter-antimatter pair annihilation, this asymmetry is found to be $\mathcal{O}(10^{-10})$ [28]. While CP-asymmetry is allowed in the SM (as discussed in Sec. 2.1.3), this is not sufficient to describe the observed matter-antimatter asymmetry by many orders of magnitude. Either there is an amount of antimatter in the universe that has been unaccounted for or some additional mechanism by which CP symmetry is violated that is beyond the SM exists.

Finally the measured mass of the Higgs boson presents an issue known as the hierarchy problem. The mass of the Higgs is sensitive to corrections from particles that couple to the Higgs field. Mass corrections from high mass particles should result in masses at much higher scales then the observed Higgs boson mass. These corrections are approximately 30 orders of magnitude larger than the measured Higgs mass and seem to require some kind of counter effect to match with observations [29]. There must be some other mechanism beyond the SM that explains this effect, with one example of a class of theories that solve this being supersymmetry.

2.2 Searches for new physics

As demonstrated in the previous section, the SM cannot be a complete model of fundamental physics. However, there is no simple extension of the SM that could naturally include the additional unexplained phenomena which has not already been ruled out by the precise measurements already made of SM processes. This leaves open a wide range of potential models and particle candidates for mediating NP and consequentially a wide range of potential methods to carry out searches for NP. Searches for NP broadly fall into two categories:

- Direct searches: Searches for the direct decay of NP particles. By carrying out a search for a decay resonance in a final state, these searches can explore the parameter space of new particle mass and relative coupling strength to the SM. This is the type of search which lead to the observation of the Higgs boson. A direct search is limited by the centre of mass energy of the collision in the collider used to perform the search. As of yet, no direct search has found evidence of NP, although they have ruled out large areas of the available parameter space for new particles.

- Indirect searches: These make use of the ability of high mass particles to be part of decays off-shell, as virtual mediating particles in the decay. As such, searches can be carried out at lower centre of mass energies then the particle being searched for. Evidence for these particles would be seen as modifications of the expected properties of SM allowed processes, or the presence of processes which are forbidden in the SM. This method requires precise predictions of the SM properties of decays and observables that are sensitive to these modifications.

2.2.1 Electroweak penguin decays

As mentioned, the sensitivity of indirect search methods is reliant on the precision of predictions for observables of SM processes and on how large a NP contribution can be with respect to the SM contribution. For this reason processes that are rare within the SM are ideal for carrying out indirect NP searches. Electroweak penguin (EWP) decays are a type of flavour-changing neutral current (FCNC) transition which cannot occur at the tree level in the SM, only at the loop level or higher. This is due to there being no coupling between the same type (down- or up-type) quarks over different generations, transitions can only take place via a charged interaction. The is forbidden at tree level and is heavily suppressed at the loop level

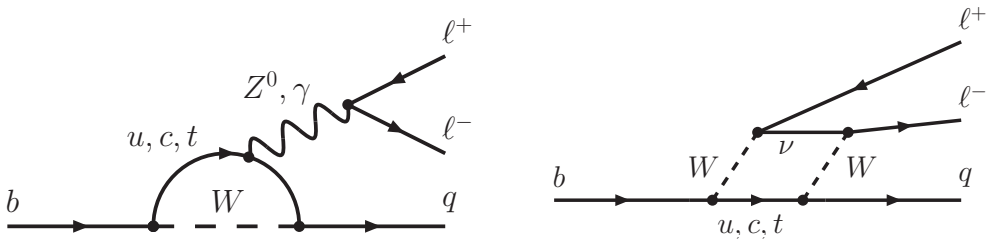

Figure 2.3: Lowest order Feynman diagrams for a $b \to q\ell\ell$ FCNC current decay. Left is the penguin diagram, right is the box diagram.

a consequence of the GIM mechanism [7]. The GIM mechanism was introduced to explain the small rate of the decay $K_S \to \mu^+\mu^-$ and required the introduction of a new fourth quark, the charm quark, in addition to the already at the time known up, down and strange quarks. This leads to a four quark unitary mixing matrix and the extension of this mixing matrix to the unitary CKM matrix preserves this lack of FCNC transition. The lack of tree level transition leads to a heavy suppression in the SM, dependent on the couplings between particles at the loop vertices. At the loop level, only the mass difference between quark generations breaks the GIM suppression, allowing for FCNC transitions. Example lowest-order Feynman diagrams for decays of this type are shown in Fig. 2.3. In this thesis, the primary process looked at involves a $b \to d$ quark level transition.

The properties of these decays can be modified by the presence of a NP particle, which both do not necessarily have the same restriction of being forbidden at tree level and can enter into the diagram at the loop level as a replacement to the allowed SM contributions. By predicting the observables of these decays in the SM and comparing them to measured values, powerful probes of the SM can be carried out.

2.2.2 Parametrising new physics

One of the most effective techniques for describing weak decays is to make use of an effective field theory which acts to separate the energy scales involved in calculations of a process. Perhaps the most famous example of this is the Fermi Theory of weak decays [30], where the short distance physics is represented by a point like vertex associated to a local operator which describes the transition. The full theory is then described by an effective theory with this local operator and the Fermi constant G_F. An example of this for β decay is shown in Fig. 2.4.

This framework allows for an effective way of carrying out SM calculations

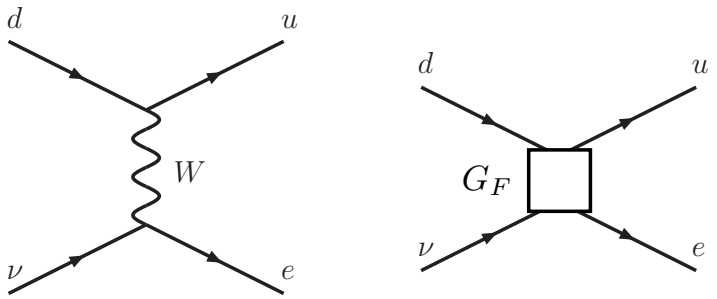

Figure 2.4: Feynman diagram for β decay in the full theory (left) and in the effective theory (right).

while allowing for a natural inclusion of NP effects. In an extended effective theory, a number of different local operators can be included for different transition types, each with an associated coupling constant. The inclusion of extra operators associated to NP transitions is then simple and leads to observables which can be measured experimentally. For an effective field theory describing decays of B mesons, this split is performed at the scale of the b mass, m_b, where the weak interactions can be treated as point-like. The couplings then describe the high energy scale and can be calculated perturbatively at the higher electroweak mass scale. The contribution of any new particle at a higher mass scale would then also act as a modification to these couplings. The operators describe the structure of the four fermion interactions. There is a clear split between the three energy scales, Λ_{QCD}, the energy scale associated with the binding of the initial and final states, m_b and $m(W)$.

A model for describing rare $b \to q$ FCNC transitions can be constructed using an effective field theory similar to the Fermi theory, where it has been generalised to all quark and lepton types. The effective Hamiltonian for a $b \to q$ quark transition (where $q = s$ or d and where $\lambda_p^{(q)} = V_{pb}V_{pq}^*$) is represented by:

$$\mathcal{H}_{\text{eff}}^{b\to q} = \frac{4G_F}{\sqrt{2}}\left(\lambda_u^{(q)}\sum_{i=1}^{2} C_i\mathcal{O}_i^u + \lambda_c^{(q)}\sum_{i=1}^{2} C_i\mathcal{O}_i^c - \lambda_t^{(q)}\sum_{i=3}^{10} C_i\mathcal{O}_i - \lambda_t^{(q)}C_\nu\mathcal{O}_\nu + \text{h.c.}\right). \quad (2.11)$$

This is constructed by carrying out a operator product expansion (OPE),

treating the weak interaction as point like at the scale of the interaction and encoding the short distance physics effects in a series of Wilson coefficients [31], C_i, of operators, $\mathcal{O}_i$, that have different Lorentz structures. The Wilson coefficients can be calculated via perturbative techniques whereas the operators describing long-distance fermion effects can require non-perturbative QCD techniques to calculate. The different operators each describe a different $b \to q$ transition and the full list of operators [32] follows, first with the current-current operators:

$$\mathcal{O}_1^p = (\bar{q}_L \gamma_\mu T^a p_L)(\bar{p}_L \gamma^\mu T^a b_L), \tag{2.12}$$

$$\mathcal{O}_2^p = (\bar{q}_L \gamma_\mu p_L)(\bar{p}_L \gamma^\mu b_L), \tag{2.13}$$

where in these operators $p = u$ or c, followed by the QCD penguin operators,

$$\mathcal{O}_3 = (\bar{q}_L \gamma_\mu b_L) \sum_p (\bar{p} \gamma^\mu p), \tag{2.14}$$

$$\mathcal{O}_4 = (\bar{q}_L \gamma_\mu T^a b_L) \sum_p (\bar{p} \gamma^\mu T^a p), \tag{2.15}$$

$$\mathcal{O}_5 = (\bar{q}_L \gamma_\mu \gamma_\nu \gamma_\rho b_L) \sum_p (\bar{p} \gamma^\mu \gamma^\nu \gamma^\rho p), \tag{2.16}$$

$$\mathcal{O}_6 = (\bar{q}_L \gamma_\mu \gamma_\nu \gamma_\rho T^a b_L) \sum_p (\bar{p} \gamma^\mu \gamma^\nu \gamma^\rho T^a p), \tag{2.17}$$

where the sum is over $p = u, d, s, c, b$, followed by the magnetic penguin operators,

$$\mathcal{O}_7 = \frac{e}{16\pi^2} m_b (\bar{q}_L \sigma^{\mu\nu} b_R) F_{\mu\nu}, \tag{2.18}$$

$$\mathcal{O}_8 = \frac{g_s}{16\pi^2} m_b (\bar{q}_L \sigma^{\mu\nu} T^a b_R) G^a_{\mu\nu}, \tag{2.19}$$

and finally followed by the semi-leptonic operators

$$\mathcal{O}_9 = \frac{e^2}{16\pi^2} (\bar{q}_L \gamma_\mu b_L) \sum_\ell (\bar{\ell} \gamma^\mu \ell), \tag{2.20}$$

$$\mathcal{O}_{10} = \frac{e^2}{16\pi^2} (\bar{q}_L \gamma_\mu b_L) \sum_\ell (\bar{\ell} \gamma^\mu \gamma_5 \ell), \tag{2.21}$$

$$\mathcal{O}_\nu = \frac{e^2}{8\pi^2} (\bar{q}_L \gamma_\mu b_L) \sum_\ell (\bar{\nu}_{\ell L} \gamma^\mu \nu_{\ell L}), \tag{2.22}$$

where the sum over ℓ is over the lepton generations e, μ or τ. Throughout these operators, the subscripts L and R refer to the chirality of the fermion (left- and right-handed respectively), T^α is the QCD colour matrix, $F_{\mu\nu}$ is the Faraday tensor

describing the photon and $G^{\alpha}_{\mu\nu}$ the tensor describing the gluons.

This procedure is very useful when carrying out NP searches as any new effects can be described as a modification of the SM Wilson coefficient for the relevant operator. It also allows the introduction of new, SM forbidden, operators with corresponding Wilson coefficients which can be measured.
For a $b \to q\ell\ell$ transition, the effective Hamiltonian simplifies to:

$$\mathcal{H}_{\text{eff}} = -\frac{4G_F}{\sqrt{2}}\lambda_t^{(q)} \sum_i (C_i\mathcal{O}_i + C_i'\mathcal{O}_i'), \tag{2.23}$$

where the operators of interest for the process are $\mathcal{O}_7$, $\mathcal{O}_9$ and $\mathcal{O}_{10}$ as well as the right-handed coupling version of each operator designated by $\mathcal{O}_i'$. There are contributions from the other operators described above, however these enter as higher-order corrections to the Hamiltonian and are generally not considering when searching for NP contributions in these decays. Note that in this simplified Hamiltonian, only the contribution involving $\lambda_t^{(q)}$ is considered. This is a consequence of the GIM mechanism, that acts to suppress FCNC decays at the loop level when there is a symmetry of the quark masses involved. As the t quark has a significantly higher mass then the other quarks, this symmetry is strongly broken and removes the suppression, causing the contribution from $\lambda_t^{(q)}$ to dominate.

The relative contributions of these operators is dependent on the kinematics of the decay modes. It is useful to describe these kinematics in terms of the the invariant mass squared of the dilepton pair, q^2. Figure 2.5 shows a schematic representation of the differential cross-section of a $b \to q\ell^+\ell^-$ transition, detailing the dominant Wilson coefficients that contribute in each q^2 region. Also shown are the contributions from $c\bar{c}$ charmonium resonances which dominate the cross-section in their respective regions. These are long distance contributions from $b \to c\bar{c}q$ transitions, where the $c\bar{c}$ resonance has decayed to a dilepton pair via a virtual photon.

The expectation value of this effective Hamiltonian can be used to derive the amplitude of a specific decay mode and thus predictions of its decay properties in the SM. The components of this expectation value divide into a coupling according to the CKM matrix, the contribution from the long distance SM-like physics described by operators and a contribution from the short distance Wilson coefficients which can receive modifications from NP. The long distance effects are decay mode dependent and include the hadronic physics between the initial and final states. The operator part describing a $b \to q\ell\ell$ transition, $B \to M\ell\ell$ where B is a B meson and M is a

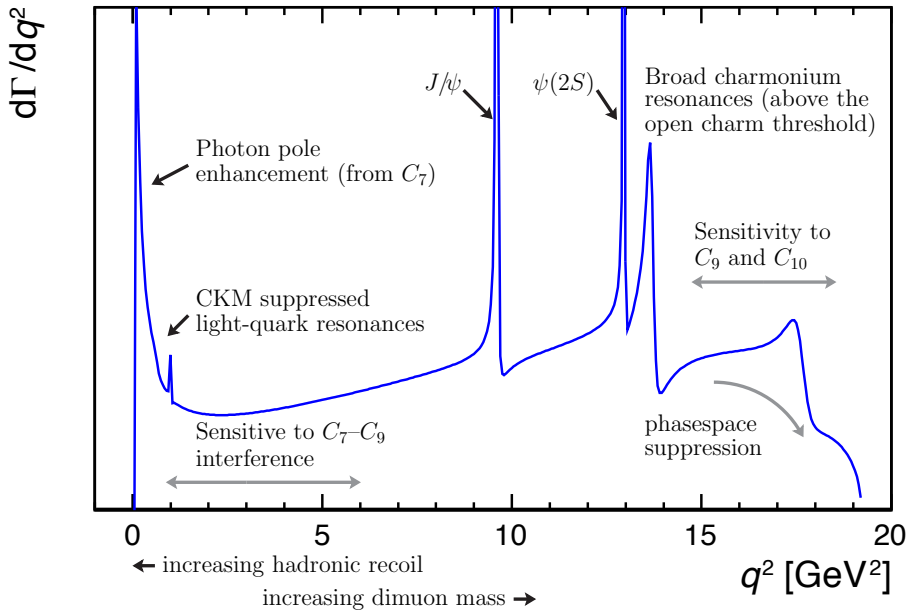

Figure 2.5: Schematic representation of the differential decay rate of a $b \to q\ell^+\ell^-$ decay, detailing the different contributions at various points in q^2. Reproduced from Ref. [32].

generic meson, can be written as

$$\langle \bar{\ell}\ell M|\mathcal{O}|B\rangle = \langle \bar{\ell}\ell|\mathcal{J}_\ell|0\rangle \cdot \langle M|\mathcal{J}_q|B\rangle + \text{non-factorisable effects} \tag{2.24}$$

$$\sim \langle \bar{\ell}\ell|\mathcal{J}_\ell|0\rangle \cdot F(q^2) + \text{non-factorisable effects} \tag{2.25}$$

where the operator $\mathcal{O}$ has been split into a leptonic current, $\mathcal{J}_\ell$, and a quark current, $\mathcal{J}_q$, and the components of the decay have been partially factorised. The hadronic part of this factorisation, $F(q^2)$, is known as a form factor and is calculable in the SM through a number of different techniques depending on the kinematic properties of the involved hadron. The additional non-factorisable terms are harder to compute. Qualitatively, they correspond to cases where the separation of the short-distance and long-distance breaks down. For example, by connecting a gluon between the short-distance loop and the spectator quark.

Form factors can be described in terms of the q^2 parameter (as shown in Fig. 2.5), where a low value of q^2 corresponds to a large hadronic recoil and vice-versa. In the low q^2 region, QCD light cone sum rules are used to calculate form factors, in the high q^2 region lattice QCD methods are used. The uncertainties on SM predictions of the properties of many EWP decays are driven by the uncertainties on these form factors calculations, with the next most significant uncertainties

being related to the non-factorisable effects. A more detailed discussion on these uncertainties is given in Section 2.3.5.

2.2.3 Observables sensitive to new physics

The key requirements to have sensitivity to NP effects are:

- A clean theoretical prediction for the observable under the SM only assumption.
- The SM contribution being small with respect to a potential NP contribution.
- An experimentally clean measurement with small or very well modelled background.

While it is not necessary for all of the above to be true to make measurements of this kind, the best opportunities to observe NP come in decay modes and observables which satisfy all of these to some degree. Electroweak penguin decays satisfy these requirements and provide access to several observables that are sensitive to NP. The observables include the differential branching fractions of the decays, CP asymmetries in the decay modes and the angular distribution of the final state particles.

Angular distributions allow access to a wide variety of sensitive variables and in particular provide sensitivity to Wilson coefficients which otherwise could not be separated from each other. How much can actually be measured is dependent on how large a data sample is available and on whether the CP states can be distinguished. Also of importance is to deliberately construct observables with reduced theoretical uncertainties. By taking into account correlations in the techniques used to make these predictions, ratios of certain angular variables can have vastly reduced theoretical uncertainties related to form factors. An example of this would be the variable P_5' [33], where at low values of q^2 the symmetries of the system can be exploited to construct an observable where the form factor contributions cancel at leading order. The full extent of how reduced these uncertainties are is still not fully agreed upon in the field and will be discussed in more detail in Sec. 2.3.5.

To fully exploit rare decays as tools for searching for NP, a wide range of measurements should be made. Different decays have contributions from different operators and allow different Wilson coefficients to be measured. For example, rare fully leptonic decays are sensitive to just C_{10} while semileptonic 4-body final state decays can be sensitive to all of C_7, C_9 and C_{10} depending on the region of dimuon invariant mass squared (q^2) the measurement is being carried out in. By taking

into account results from a number of different decay modes in a consistent way, as provided by the effective field theory model and as further discussed in Sec. 2.3.6, some of the most sensitive indirect searches for NP with the current generation of high energy particle physics experiments can be carried out.

2.3 Current experimental status

In the following section, an overview of some of the recent results in rare decays is given and as well as some of the future prospects for these measurements.

2.3.1 Very rare leptonic decays

The purely leptonic rare decays $B^0_{(s)} \to \mu^+\mu^-$ can act as very powerful probes of the SM, mainly due to the precision of theoretical predictions associated with the mode. This is mainly due to there being a lack of a hadronic form-factor required in the calculation, leading to small theoretical uncertainties. The SM branching fraction of these decays has been calculated with an uncertainty of approximately 10% for both [34]. They are also some of the most heavily suppressed modes in the SM due to CKM and helicity suppression and the experimental signature for the modes of just two muons is very clean and can be identified with very little background. The first observation of the modes has been made in a joint analysis by LHCb and CMS [35] with the current best measurement of the branching fraction and first single experiment observation made by LHCb [36] as well as measurements performed by ATLAS [37]. The measurements of the branching fraction made so far are consistent with SM predictions and provide strong constraints on the coefficient C_{10} in any NP scenario.

2.3.2 Radiative decays

Radiative decays of the type $b \to s\gamma$ are governed by the photon penguin operator and provide access to the coefficient C_7 as well as access to the chirality flipped version of the operator C_7'. In the SM, the photons emitted are predominantly left-handed, with a small right-handed component which is suppressed by the ratio of s and b quark masses. By measuring the photon polarisation in these decays, the degree of this suppression can be measured and compared to predictions as a test of the SM. Time-dependent techniques have been used to extract photon polarisation parameters from the time dependent CP-asymmetry of B^0 and $\overline{B}^0$ decays in the modes $B^0 \to K^{*0}\gamma$ [38, 39] and $B^0 \to K^0_S\rho\gamma$ [40, 41] and have shown no evidence for right-handed polarised photons. Photon polarisation has also been measured in the decay mode $B^0_s \to \phi\gamma$ [42], where a fit of the decay time distribution of the decay allows the polarisation to be extracted. In this mode, the results of this extraction are consistent with SM predictions. Other radiative decays that have been measured including the first observation of photon polarisation in a $b \to s\gamma$ transition in the decay $B^+ \to K^+\pi^-\pi^+\gamma$ [43] and an angular analysis of the decay

$B^0 \to K^{*0} e^+ e^-$ [44] at low values of q^2 which provides access to photon polarisation sensitive observables. These are all also consistent with SM predictions and put constraints on any NP value of $C_7^{(\prime)}$.

2.3.3 Anomalies in $b \to s l^+ l^-$ decays

Electroweak penguin decays, which proceed via $b \to s\ell^+\ell^-$ transitions, provide additional sensitivity to the Wilson coefficient C_9. There have been a number of recent results in rare electroweak penguin decays which proceed via $b \to s\ell^+\ell^-$ transitions which are beginning to show tensions with the SM. The differential branching fractions of the $b \to s\ell^+\ell^-$ decays $B^+ \to K^+\mu^+\mu^-$ [45], $B^0 \to K^0\mu^+\mu^-$ [45], $B^+ \to K^{*+}\mu^+\mu-$ [45], $B^0 \to K^{*0}\mu^+\mu^-$ [46] and $B_s^0 \to \phi\mu^+\mu^-$ [47] have all be found to be systematically lower than SM predictions. A comparison of the measured differential branching fractions and the SM predictions of the differential branching fractions are shown in Fig. 2.6.

As previously mentioned, angular analyses provide access to a number of additional observables which can be predicted in the SM and then measured. Perhaps the most high profile of these analyses is the analysis of the decay $B^0 \to K^{*0}\mu^+\mu^-$. Measurements of optimised observables [33] with reduced (see Sec. 2.3.5 for a more detailed discussion) theoretical uncertainties have been made. By using observables constructed as ratios of angular coefficients, the correlations between from factors from the different operator contributions can be taken into account, cancelling out many of the hadronic uncertainties. The LHCb measurement [53] of the observable

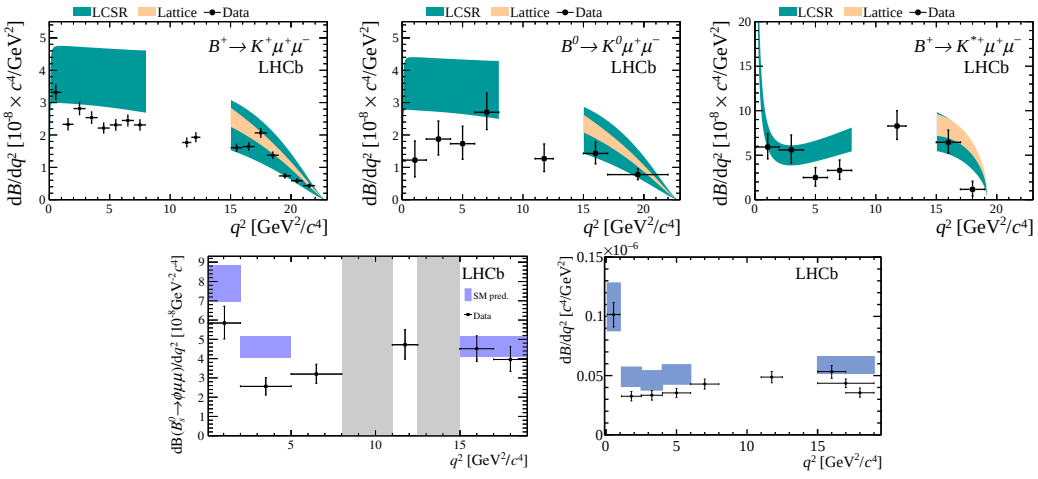

Figure 2.6: Differential branching fractions for the decays $B^+ \to K^+\mu^+\mu^-$, $B^0 \to K^0\mu^+\mu^-$, $B^+ \to K^{*+}\mu^+\mu^-$ [45], $B_s^0 \to \phi\mu^+\mu^-$ [47] and $B^0 \to K^{*0}\mu^+\mu^-$ [46] as measured by the LHCb experiment as well as SM predictions from Refs. [48–52].

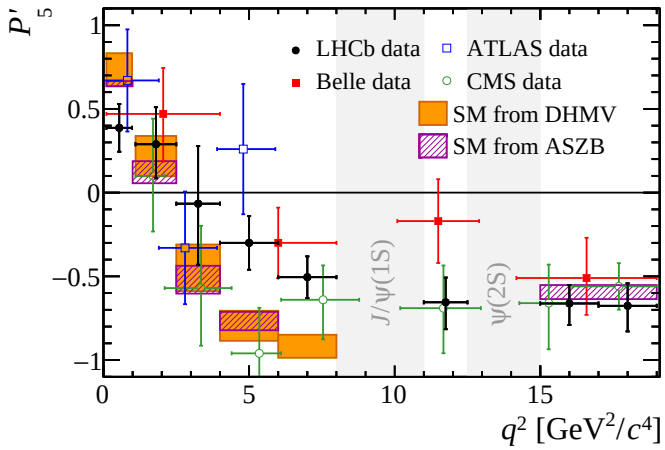

Figure 2.7: Measurement of P'_5 in bins of q^2 from LHCb [53], ATLAS [54], CMS [56] and Belle [55] as well as SM predictions from Refs. [52, 57].

P'_5 continues to show tensions with respect to the SM predictions, with a combined tension of 3.4 standard deviations over two bins of q^2. Figure 2.7 shows the result of the analysis as well as results from ATLAS [54], Belle [55] and CMS [56] along with SM predictions [52, 57].

2.3.4 Lepton flavour universality

Further hints of new physics are found in measurements investigating lepton flavour universality (LFU) between $b \to se^+e^-$ and $b \to s\mu^+\mu^-$ decays. In these decays, the coupling of the electrons and the muons to the mediating particle in the electroweak penguin decays should be identical, excluding small corrections due the difference in masses. A verified observation outside the small theoretical uncertainties is a clear indication of NP. Figure 2.8 shows the experimental measurements of R_K and $R_{K^{*0}}$ from LHCb [58, 59], BaBar [60] and Belle [61]. The values of these ratios should be unity in the SM. Both LHCb measurements show a discrepancy with respect to the SM in the same direction at the level of 2.6 standard deviations for R_K and 2.4 and 2.5 standard deviations in the low and central q^2 regions respectively for $R_{K^{*0}}$. While neither of these measurements are individually at the significance to be considered evidence for NP, the presence of similar discrepancies in two $b \to s\ell^+\ell^-$transitions which may indicate a departure from LFU are some of the most intriguing hints of NP seen to date.

The measurements of R_K and $R_{K^{*0}}$ are not the only hints for lepton non-universal couplings. Measurements of the semi-leptonic tree level decays modes of the type $b \to c\ell\nu_\ell$ also show deviations from SM predictions. In this case, the

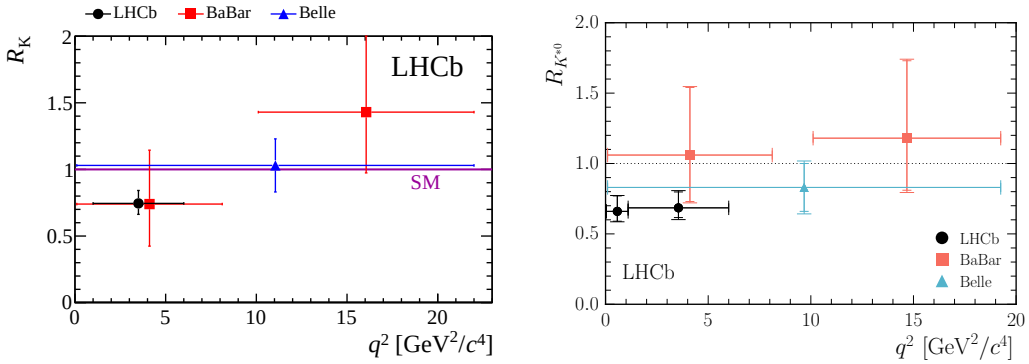

Figure 2.8: Measurements of R_K and $R_{K^{*0}}$ in varying regions of dilepton invariant mass squared (q^2) from LHCb [58, 59], BaBar [60] and Belle [61]. Also shown is the SM expectations of unity for both modes.

predictions of the ratios of these branching fractions are not unity due to the much larger lepton mass difference, however they are still very precisely predicted. The ratios R_D and R_{D^*} have been measured by BaBar [62, 63] and Belle [64–66] as well as measurements of R_{D^*} by LHCb [67, 68]. When the combination of these measurements is taken into account along with the correlation between R_D and R_{D^*}, the difference between the measured and SM values is approximately 3.8 standard deviations [69].

2.3.5 Theoretical uncertainties

The tensions in the above measurements are clearly of great interest, however there is still some discussion in the theory community on whether these measurements are truly indicative of NP or are a result of underestimated QCD effects [70–72]. Experimentally, the main $c\bar{c}$ resonances (the J/ψ and the $\psi(2S)$) are vetoed in analyses and theoretical predictions are carried out outside of the q^2 region dominated by these resonances. Figure 2.5 shows a schematic representation of this. However, while the regions with resonances are very efficiently vetoed, it is not necessarily true that hadronic effects are small outside of these veto regions. The non-factorisable hadronic contributions to the operators is also q^2 dependent and it is possible that the uncertainty related to these effects over the whole q^2 has been underestimated.

One way of resolving this issue is to make an experimental measurement of this interference. This is done be carrying out a full amplitude analysis of the entire q^2 range. This has already been carried out by LHCb for the decay $B^+ \to K^+\mu^+\mu^-$ [73]. No large interference is found in this case, but this does not mean it is true for the vector rather then scalar meson case. The same measurement but

carried out for $B^0 \to K^{*0}\mu^+\mu^-$ will be of great interest in the future but will be significantly more experimentally challenging to carry out. Methods by which this could be carried out have been suggested in Refs. [74, 75]. It is important to note that while differential branching fractions and angular observables could suffer from these underestimated uncertainties, measurements related to LFU do not. As such, there is still good indication of possible NP effects in this area.

2.3.6 Global fits

No single measurement has yet to unambiguously show data inconsistent with the SM-only hypothesis, however there does appear to be a pattern of similar tensions emerging in $b \to s\ell^+\ell^-$ decays. By making use of the OPE formalism and converting the measured tensions into a corresponding NP Wilson coefficient, global fits of the measurements can be carried out. This results in the preferred value of these NP coefficients to consistently describe the tensions. Figure 2.9 shows an example of one such fit, including inputs from angular analyses of $b \to s\ell^+\ell^-$ transitions and the latest results relating to LFU. In this fit, the resulting preferred fit value of the vector and axial-vector coupling coefficients is approximately 4 standard deviations away from the SM values. This can be explained by a new left-handed NP current in $b \to s\ell^+\ell^-$ decays.

Of particular interest in these fits is that both the set of measurements related to the flavour anomalies in the muon sector (*e.g.* differential branching fractions and angular analyses) and the set related to lepton flavour universality point towards a SM modification of the same Wilson coefficients. Taking all measurements into account, the indications of NP can become greater than 5 standard deviations according to a number of different global analyses [76–80].

2.3.7 Future prospects for new physics searches

The anomalies currently being observed in the field of flavour physics are clearly of great interest in the search for NP. Further results that can either confirm or rule out these anomalies are highly anticipated and in the case that they are confirmed, the resulting tests of the structure of NP will be extremely important. In the short term, updates of many of the measurements detailed above will be needed before any single measurement can exceed the level of 5 standard deviations. In addition, many additional measurements related to lepton universality in the $b \to s\ell^+\ell^-$ sector are underway by a number of different collaborations. It should also be possible to experimentally assess the hadronic interference of charm loops to the non-resonant

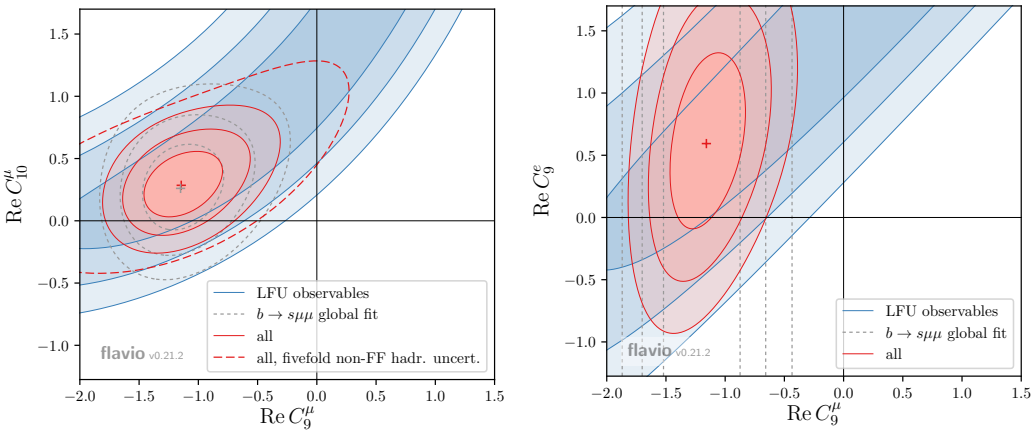

Figure 2.9: Global fit including inputs from LFU measurements and $b \to s\mu\mu$ from multiple experiments for C_9/C_{10} and C_9/C_9'. The origin of these plots indicate the SM value of these coefficients. Reproduced from Ref. [77].

modes, this should help clarify the question of whether some of anomalies already observed are due to underestimated theoretical uncertainties or not.

If NP contributions are confirmed, the flavour structure of this NP will be of great interest. One of the main ways of probing any potential structure would be to expand the series of measurements into different flavour transitions. The available datasets at LHCb are reaching the point where more detailed analyses of $b \to d\ell^+\ell^-$ transitions can be carried out. This will include angular analyses and tests of LFU in the $b \to d\ell^+\ell^-$ partners of the already investigated $b \to s\ell^+\ell^-$ transitions. On their own, these decays will have excellent sensitivity in searches for NP for the same reasons that $b \to s\ell^+\ell^-$ transitions do. In combination with $b \to s\ell^+\ell^-$ measurements, they will give an indication of whether any NP effect is limited to the $b \to s\ell^+\ell^-$ sector or if there is a more rich flavour structure. Searches for $b \to d\ell^+\ell^-$ transitions with the present datasets are vital in setting the groundwork for these analyses in the future. It is also possible to exploit the fact that the dominant contribution to the difference in rates of $b \to d\ell^+\ell^-$ and $b \to s\ell^+\ell^-$ transitions is the ratio of CKM factors, $|V_{td}/V_{ts}|$. By measuring the decay rates of partner modes and taking into account the additional differences in hadronic contributions, this ratio can be extracted from data. A measurement of this kind has been carried out using the ratio of branching fractions of the decay modes $B^+ \to K^+\mu^+\mu^-$ and $B^+ \to \pi^+\mu^+\mu^-$ [81] and more measurements of this kind will be able to contribute to the world average measurement of the ratio.

Finally, the presence of any indication of NP measured due to the indirect methods detailed above would motivate a new generation of direct searches. Al-

though there has so far been no NP found in direct searches being carried out at the LHC, results from indirect searches and the corresponding theoretical models that are built to describe these results could provide indications of where direct searches will be most effective. If this indicates regions outside of the energy range of the current machines and experiments this could motivate the next generation of high energy physics experiments.

Chapter 3

The LHCb experiment

The following chapter provides details of the experiment at which the data presented in this thesis were collected. A brief description of the Large Hadron Collider [82] is given with a focus on the machines suitability to be used in studies of heavy flavour physics. This is followed by a detailed description of the LHCb [83] detector, describing the individual components of the detector and how data are collected from the detector in both Run 1 and Run 2 of the LHC. Finally, a short overview of planned future upgrades of the LHCb detector is presented.

3.1 The Large Hadron Collider

The LHC is a proton–proton collider with an approximate circumference of 27 km located at the European Organisation for Nuclear Research (CERN). It is, at the time of writing, the worlds largest and most powerful particle accelerator, designed to study physics at centre-of-mass (c.o.m.) collision energies of 14 TeV. The ultimate aim of the LHC is to discover physics beyond the SM.

The LHC makes use of the existing accelerator complex at CERN from previous particle accelerators to progressively increase the energy of the protons up to the operational energy. Protons are produced from hydrogen gas and injected into the LHC through the injector chain comprising, in the order that the protons are accelerated through, the Linear Accelerator (LINAC), the Proton Synchrotron Booster (PSB), the Proton Synchrotron (PS) and the Super Proton Synchrotron (SPS). The LHC can also operate with ion beams, either colliding two lead ion beams together or one lead ion beam with a proton beam. The work detailed in this thesis is based only on data collected from proton-proton collisions and only the details related to these collisions will be presented. The CERN accelerator complex, as well as the

location of the main LHC experiments is illustrated in Fig. 3.1.

The proton beams are accelerated and contained within the LHC ring with a combination of radio frequency (RF) cavities and superconducting magnets, respectively. The LHC ring is made up of eight sectors, with each sector containing a straight section used for accelerating the beam and an arc section used to bend the beam. The straight sections contain the RF cavities which boost the energies of the two proton beams and well as the interaction points at which the beams are brought together to collide. The arc sections are made of superconducting dipole magnets with field strengths of over $8T$, with this field being required to contain the beam at the design energies of 7 TeV each. Quadrupole, sextupole and octupole magnets are used throughout the machine to focus the beam and allow for the collision of the beams.

Proton beams are made ready for collisions in a multistage process involving the magnets in the machine. First the beams are accelerated by the injector chain up to an energy of 450 GeV before being injected as two separate beams which counter-rotate in the LHC. The machine then enters the ramping stage, where they are accelerated by the RF cavities in the main LHC ring. Upon reaching the desired energy per beam the machine enters the squeezing stage, whereby the focussing magnets are used to shape the beam into the optimal conditions for collisions. Finally, after the beams have been squeezed, "stable beam" conditions are achieved, allowing for beam collisions and data collection by the experiments at the LHC.

The proton beam itself is made up of bunches of protons, with the separation between the bunches in time defining the operational frequency of the machine. The design bunch separation is 25 ns, corresponding to a crossing frequency of 40 MHz. The collision rate of protons is defined by the luminosity of the beam, $\mathcal{L}$, which is calculated as [84]

$$\mathcal{L} = \frac{N_B (N_p)^2 f}{4\pi\sigma^2} \cdot S, \tag{3.1}$$

where N_p is the number of protons in each bunch, N_B is the number of bunches in each beam, f is the revolution frequency of the machine, σ is the transverse beam size and S is a luminosity reduction factor resulting from the beam crossing angle. Each proton bunch has a maximum of 1.15×10^{11} protons, with this number falling off over the course of a fill of the LHC as collisions occur. The total number of bunches in the beam varies over the course of machine operations and the bunch scheme is one of the key choices in determining the rate at which collisions occur.

There are four main particle detectors at the LHC, each one being located on one of the straight sections of the ring. The first two are high luminosity ex-

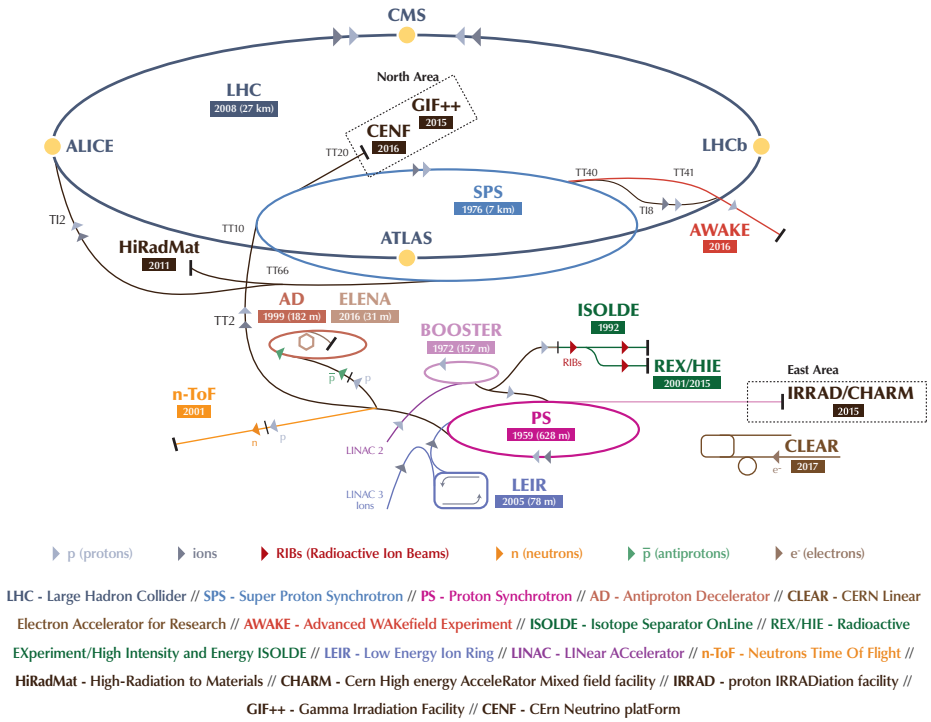

Figure 3.1: Schematic representation (not to scale) of the CERN accelerator complex and LHC injector chain. The accelerators that are relevant to the operation of the LHC are LINAC 2, the PS Booster, the PS and the SPS. Reproduced from Ref. [85]

periments designed primarily to perform direct searches for high energy particles, the ATLAS experiment and the CMS experiment, which are often referred to as general purpose detectors (GPDs). The design peak luminosity of the LHC which the GPDs operate at is $\mathcal{L} = 1 \times 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. The other two detectors, ALICE and LHCb, are more specialised experiments that operate at lower luminosities. The ALICE detector is designed to study lead–lead and lead–proton collisions with a design peak luminosity of $\mathcal{L} = 2 \times 10^{29}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. The LHCb detector is designed to make precision measurements of heavy flavour decays and has a design operational luminosity of $\mathcal{L} = 2 \times 10^{32}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. A more detailed overview of the LHCb detector, its operational luminosity and how the detector interfaces with the LHC is given in Sec. 3.2.

Although not the primary motivation for its construction, the LHC provides an excellent opportunity to investigate heavy flavour physics due to the large $b\bar{b}$ production cross-section at the machine. The production of heavy quark-antiquark pairs, here referred to as $Q(\bar{Q})$, are described in QCD by a number of processes at different orders in the strong coupling constant, α_S [86]. The leading order processes

are given by the quark-antiquark annihilation and gluon-gluon fusion processes

$$q + \bar{q} \rightarrow Q + \bar{Q}, \tag{3.2}$$

$$g + g \rightarrow Q + \bar{Q}. \tag{3.3}$$

At the high energies of operations of the LHC, these leading order processes receive corrections from and are dominated by the next-to-leading order processes

$$q + \bar{q} \rightarrow Q + \bar{Q} + g, \tag{3.4}$$

$$g + q(\bar{q}) \rightarrow Q + \bar{Q} + q(\bar{q}), \tag{3.5}$$

$$g + g \rightarrow Q + \bar{Q} + g. \tag{3.6}$$

The high energy behaviour of heavy flavour production from pp collisions is primarily described by the flavour excitation and gluon splitting production mechanisms which contribute as part of the third next-to-leading process above [87]. The total cross-section to produce b-hadrons has been measured to be approximately $295\,\mu\text{b}$ at c.o.m. $7\,\text{TeV}$ and $600\,\mu\text{b}$ at c.o.m. $13\,\text{TeV}$ [88]. This b production cross-section includes not just a large number of B mesons but also a large number of b-baryons, providing a wide variety of potential decay channels to explore in the high energy proton-proton collision environment.

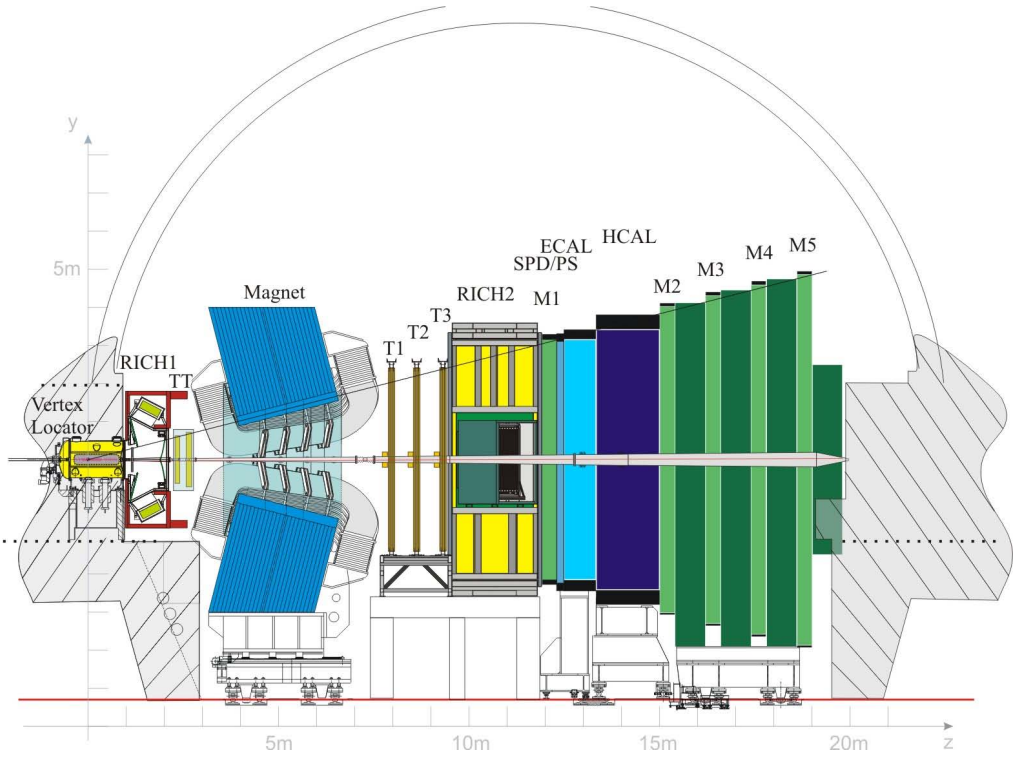

Figure 3.2: A cross section of the LHCb detector, as viewed from the side. The z axis indicates the direction of the beam line [83].

3.2 The LHCb detector

The LHCb detector is a single-arm forward spectrometer designed to take advantage of the large $b\bar{b}$ production cross-section at the LHC. A schematic of the detector is shown in Fig. 3.2. At high centre of mass energies, b and $\bar{b}$ hadron pairs are produced in a localised cone in the forward or backwards direction with respect to the interaction point. Ideally a detector could be constructed in both the forward and backward region. However, due to monetary and spatial limitations, the LHCb detector is constructed in just the forward region. In this way, the geometry of detector is optimised to make measurements of heavy flavour hadrons.

The coordinate system of the detector is defined around the collision point of the proton beams in the detector. The z-axis of the detector is aligned with the beam line, with positive z in the direction from the collision point to the dipole magnet. The x-axis is in the horizontal direction with respect to the beam line and the y-axis in the vertical direction. The angular acceptance of the detector is approximately 15 to 300 (250) mrad in the x (y) direction. This is equivalent to a

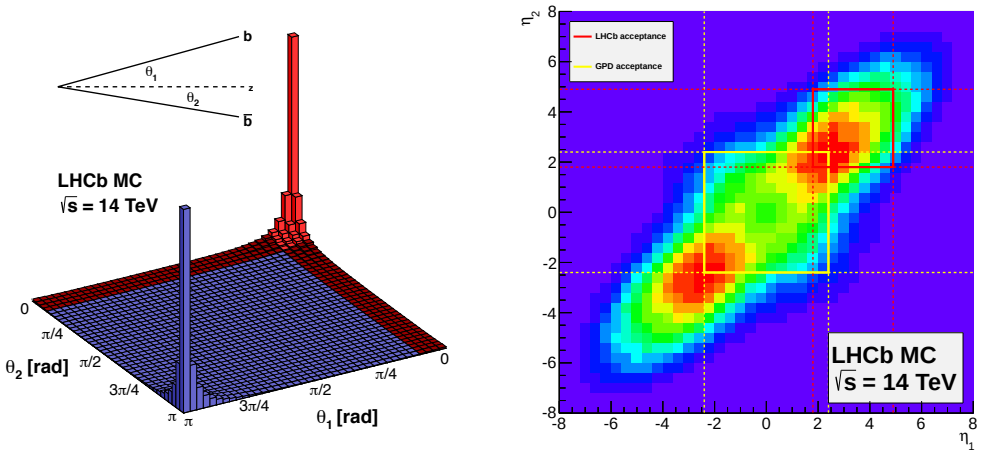

Figure 3.3: Distribution of $b\bar{b}$ production in terms of angle with respect to the beam (z) axis (left) and with respect to the pseudorapidities (η) of the $b\bar{b}$ pairs (right). The area in red indicates the LHCb angular acceptance, while the area in yellow indicates the general purpose detector acceptances. Reproduced from Ref. [89].

pseudorapidity range of $2 < \eta < 5$ with pseudorapidity being defined as

$$\eta = -\ln(\tan(\theta/2)), \tag{3.7}$$

where θ is the angle with respect to the beam-line. The lower limit of the detector angular acceptance is due to the presence of the beam-pipe, the upper limit is based on the wedge geometry of the detector. There is a higher upper acceptance limit in the horizontal direction as this is the bending plane of the LHCb dipole magnet, discussed further in Sec. 3.2.4. The angular distribution of $b\bar{b}$ pairs as well as the acceptance of LHCb and the GPDs is shown in Fig. 3.3. Although the choice of the acceptance of the LHCb detector was motivated by the $b\bar{b}$ production in the forward region, it is also advantageous that the detector covers a complementary region to the GPDs. The LHCb detector is capable of making a variety of measurements in the forward-region and is not limited to heavy flavour physics.

As previously mentioned, the LHCb detector is designed to operate at lower luminosities than the design luminosity of the LHC itself. This is motivated by the requirement of a heavy flavour physics experiment requiring excellent vertexing precision of decays in the detector. At high luminosities, multiple interactions between the beams occur within a single bunch crossing, with the average number of interactions visible to the detector known as the pile-up. At higher pile-up it becomes increasingly difficult to associate B meson decays to a particular production point.

Track reconstruction and particle identification also become more challenging. To reduce the impact of these problems, the luminosity is reduced to reduce the pile-up. The luminosity of the beam at the detector is reduced from the running luminosity of the LHC by offsetting the colliding beams. This offset can be modified over time, maintaining a constant operation luminosity throughout each fill, in contrast to ATLAS and CMS which operate at much higher luminosity which fall off over time and the number of protons in each bunch reduces. The design luminosity for LHCb is $\mathcal{L} = 2 \times 10^{32}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ however over the course of Run 1 the operational luminosity has been higher at $\mathcal{L} = 4 \times 10^{32}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$.

There are two main requirements for performing high quality flavour physics measurements that are considered in the design of the LHCb detecter: first the need to be able to accurately measure the positions of the primary vertex (PV) and secondary decay vertices of the pp interactions, as a displaced secondary vertex is a key signature of a heavy flavour decay; and second the need for high quality particle identification (PID) to distinguish between topologically similar decays that differ only by final state particle type. The detector tracking and particle identification systems and how they fulfil these requirements will be explained in detail in the following sections.

3.2.1 Tracking

The tracking system comprises a high precision vertex locator system (VELO) and a series of tracking stations located both up- (TT) and downstream (T1 – T3) of the detectors dipole magnet. The combination of the individual tracking stations allows for high precision measurements of track location to be made throughout the detector, with the emphasis being on precision in the region close to the interaction point. The inclusion of a dipole magnet with a large bending power of also allows for high precision measurements of the momenta of individual tracks, essential for good resolution of reconstructed invariant masses.

One of the key requirements of the experiment is the ability to make precise position measurements of primary and secondary vertices of the relevant particle decays close to the interaction point. The displaced secondary vertices are key features of heavy flavour decays and identifying these vertices is essential in distinguishing signal candidates from background. At the LHC energies, the lifetime of a B meson leads to a decay length of approximately 1 cm, good resolution with respect to this distance is needed to identify these vertices. Precise measurements of the decay length are also essential in a number of measurements, with the most stringent requirement being the capability to resolve the fast oscillations of the B_s^0 resulting

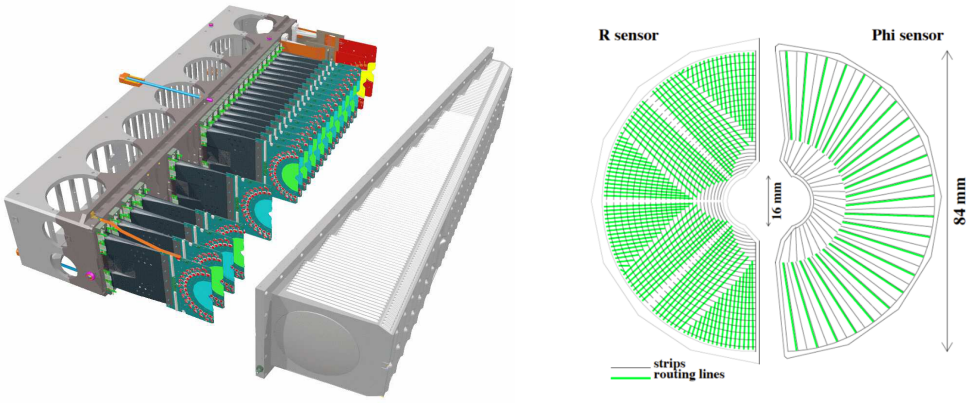

Figure 3.4: Simulated view of one side of the VELO detector (left), showing the module support, modules and RF box [83] and a schematic of VELO sensors (right), showing locations of sensor strips on both the R and Φ sensors [90].

from $B_s^0 - \overline{B}_s^0$ mixing which have a mixing frequency of $17.7\,\mathrm{ps}^{-1}$.

3.2.2 VELO

The VELO is designed to make precise track measurements in the region around the interaction point. It consists of a series of retractable modules around the beam comprising silicon micro-strip sensors providing radial and azimuthal coordinate measurements, the R and Φ sensors respectively. There are 21 modules on each side of the VELO detector, as well as an additional module on each side with only R sensors, known as the pile-up system.

Each half of the detector is contained within an aluminium RF-box, which is held under a separate vacuum from the main LHC machine. The purpose of this box is to provide shielding for the detector against RF pickup from the beam as well as protecting the LHC vacuum from outgassing of the detector. A representation of one side of the VELO detector as well as a schematic of both a R and Φ sensor is shown in Fig. 3.4.

As previously discussed, good identification of the primary and secondary vertices of a heavy flavour decay is a key feature used to separate the decays from background decays. To enable high precision measurements of these vertex locations, the sensors of the VELO detector are placed within 7 mm of the beam. However, during the injection, ramping and squeezing stages required to ready the proton-proton beams for collision, the position and profiles of the beams is not the same with respect to stable beam conditions in which the data collection is carried out. This necessitates that the two halves of the VELO are retractable, and that the

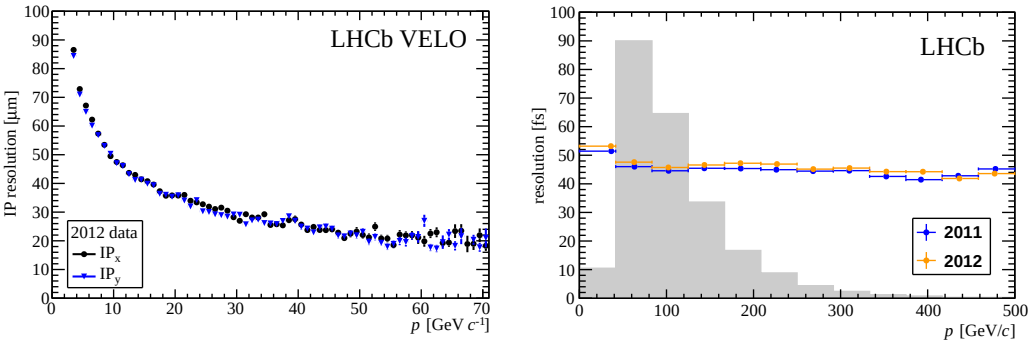

Figure 3.5: VELO impact parameter resolution in both the x and y direction (left) and decay time resolution (right) as a function of track momenta as measured from LHCb Run 1 data [90]. Note that the shaded region in the right plot shows the distribution of momenta of charged particles produced in the detector on an arbitrary scale.

modules are moved into position each time stable beam are declared by the LHC control centre.

Performance of the VELO in Run 1 has been excellent, with good resolution on track location and momenta achieved by the detector. One key parameter of the tracks of heavy flavour decay products is the impact parameter (IP), defined as the distance between the PV of the event and the tracks closest point to the PV. Decay products that result from a heavy flavour decay will be displaced from the PV and so have a larger IP than background events. The IP resolution as a function of track momenta is shown in the left plot of Fig. 3.5 as measured from 2012 data. The best resolution is achieved at high momenta. There is a limit to the resolution that can be obtained even at high momenta, mainly driven by multiple scattering effects of particles passing through the material of the VELO. The decay time resolution has also been measured in both 2011 and 2012 data as a function of track momenta and this is shown in the right plot of Fig. 3.5. Here the momenta dependence on the decay time resolution is not as a large a component as for the IP resolution, mainly due to the the secondary vertex resolution of the VELO being the primary source of uncertainty. It can be seen from this plot that the typical decay time resolution as measured in Run 1 is approximately 50 fs, much smaller than the previously stated B_s^0 oscillation frequency of 17.7 ps^{-1}.

3.2.3 Tracking stations

In addition to the VELO, there are four planar tracking stations positioned throughout the detector; the Tracker Turicensis (TT) which is located upstream of the detector magnet (described in Sec. 3.2.4) and the T1 – T3 trackers located downstream

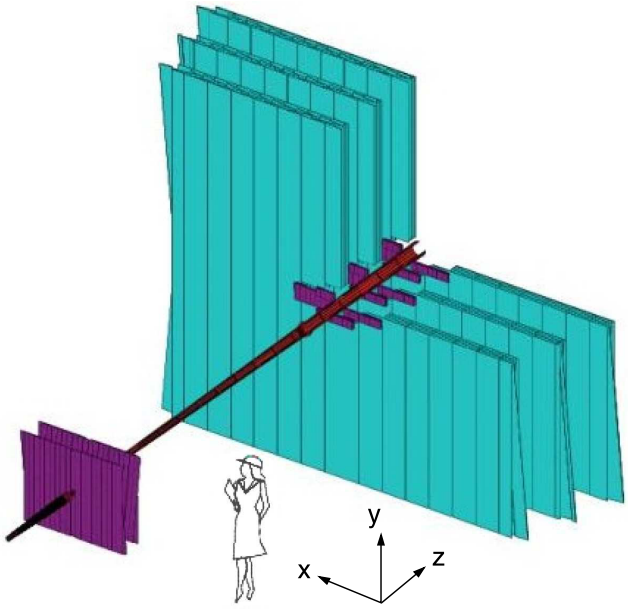

Figure 3.6: Layout of the TT, OT and IT tracking stations in the LHCb detector [83]. Purple regions indicate the TT (left of plot) and IT (right of plot) and the blue regions indicate the OT.

of the magnet. The inner sections of the T1 – T3 trackers make up the Inner Tracker (IT) of the experiment and the outer sections make up the Outer Tracker (OT). The TT and IT consist of silicon microstrip detectors, whereas the OT is a straw tube detector [91]. Figure 3.6 shows the layout and relative sizes of the TT, OT and IT.

The TT detector covers the full angular acceptance of the experiment, with an active detection area of around $8.4\,\mathrm{m}^2$. It is made up of 4 layers of silicon microstrip sensors, with the first and last layers arranged such that the strips are placed vertically and the second and third layers placed such that the strips are $\pm 5^\circ$ with respect to vertical respectively, allowing for measurements of track vertical positions. The silicon microstrip sensors are arranged so as to give a track hit resolution of approximately $50\,\mu\mathrm{m}$. The 4 layers are treated as two pairs of layers, where the separation between the two pairs is larger then between the layers themselves. This division aids in the performance of track reconstruction algorithms. The top left image in Fig. 3.7 shows the layout of the silicon sensors within a single TT layer.

The IT is laid out in a similar way to the TT, with each station of the IT having four layers, the second two of with are placed $\pm 5^\circ$ with respect to vertical. In addition, the silicon sensors are arranged to have approximately the same track resolution as the TT. The IT has a relatively smaller area then the TT, with an active detection area of around $0.3\,\mathrm{m}^2$. This area is chosen so as to limit the occupancies

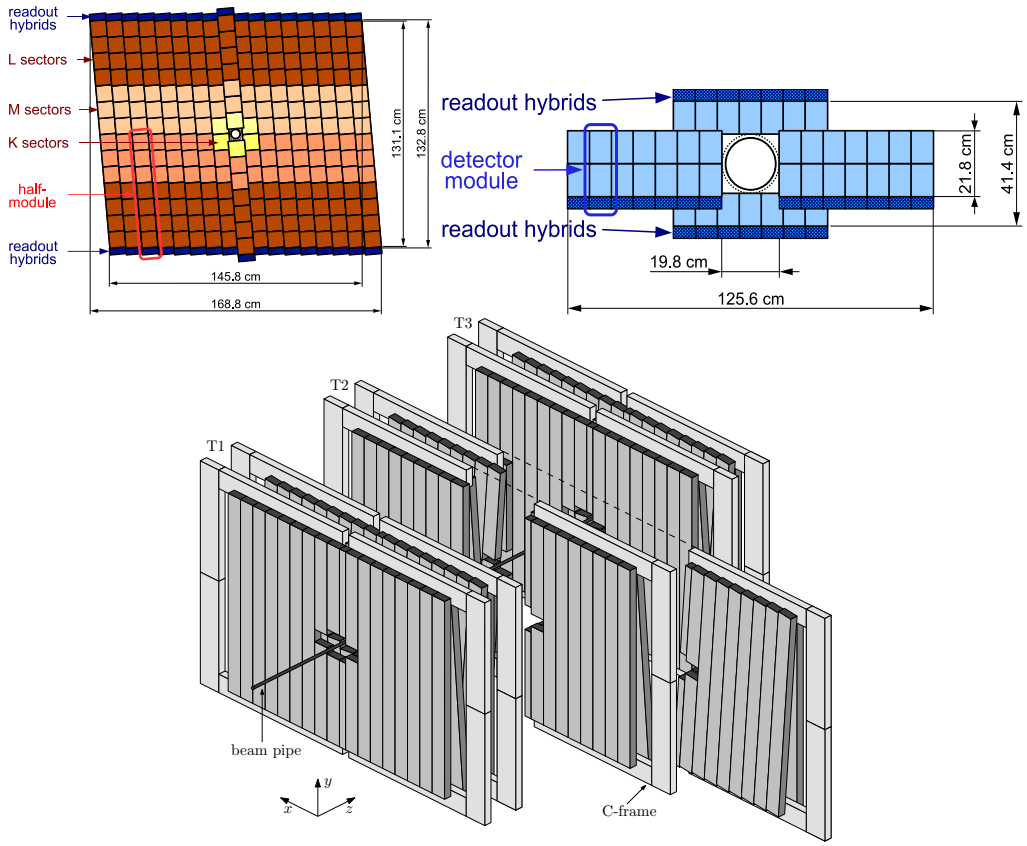

Figure 3.7: Top - layouts of a individual TT module (left) and IT module [83] (right). Bottom - schematic of the OT straw-tube modules, showing layer and station arrangement as in the detector [91].

of the OT straws to 10% in nominal running conditions.

A different detector technology is used for the OT, the primary reason being cost of covering the much larger active detection area needed to cover the full angular acceptance at these more downstream stations. The OT is a gaseous straw tube detector with an active detection area of around $30.0\,\mathrm{m}^2$. Charged particles passing through the tube ionise the gas within the tubes, with the time taken for the charge to be collected by an anode in the centre of the tube (the drift time) giving a measurement of the track position. Once again there are four detection layers for each station of the OT, with the same layout of two layers with detector components placed vertically and two layers where the components are placed at 5° with respect to vertical.

The performance of the TT tracker has been measured in Run 1 data by investigating the track resolution from daughter tracks of a clean sample of the decay $J/\psi \to \mu^+\mu^-$. By comparing the measured hit position in a given TT layer to the

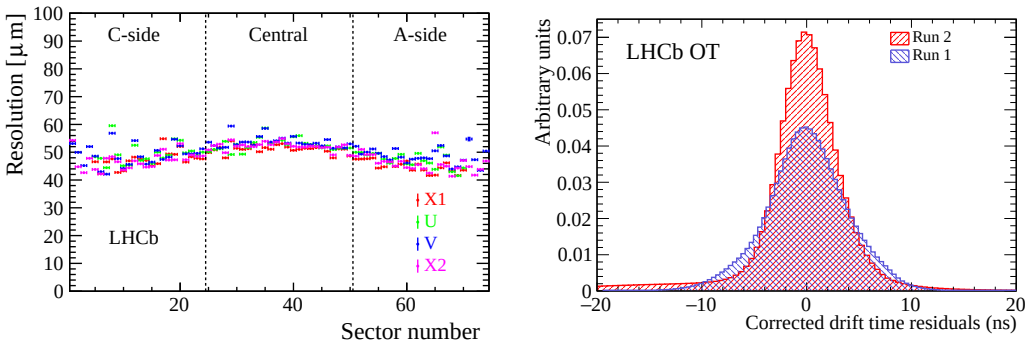

Figure 3.8: Left - Hit resolution across all modules of the TT, where X1 and X2 refer to the vertical TT modules and V and U refer to those placed at a 5° angle to the vertical position [93]. Right - Drift time residual distribution for both Run 1 and Run 2 data [92].

extrapolated hit position from all other hits associated to the track, its resolution can be determined. This resolution across the entire module for the TT is shown for 2012 data in the left plot of Fig. 3.8, where the average resolution is found to agree with the design expectations. Outer regions have better resolution due to increased charge sharing between silicon strips at greater angles with respect to the beam-pipe. Similar measurements has been performed for the IT and the resulting resolution is also in agreement with design expectations.

The performance of the OT has been measured for both Run 1 and Run 2 in Ref. [92], in particular the drift time resolution (and thus the resulting track position resolution). The distribution of drift times and thus the resolution is shown in the right plot of Fig. 3.8. Between Run 1 and Run 2, the alignment calculation and calibration procedure has been improved. This leads to an approximately 20% improvement in the drift time resolution for Run 2.

3.2.4 Magnet

A dipole magnet with an approximate bending power of 4 Tm is used in combination with the tracking systems to make measurements of the momenta of charged particles. The magnetic field covers the full acceptance of the detector, allowing the tracking systems to make momenta measurements for any particle that has not been swept out of the acceptance by this point. Typically, particles with $p \lesssim 1\,\text{GeV}/c$ will be swept out by the magnetic field. The VELO and TT provide measurement points on tracks before the magnet and the IT/OT provide measurements after the magnet. Note that the field within the VELO itself is small, resulting in straight tracks within the VELO and allowing for the best precision in vertex measurements. The RICH detectors are shielded using iron shielding to reduce the stray field from

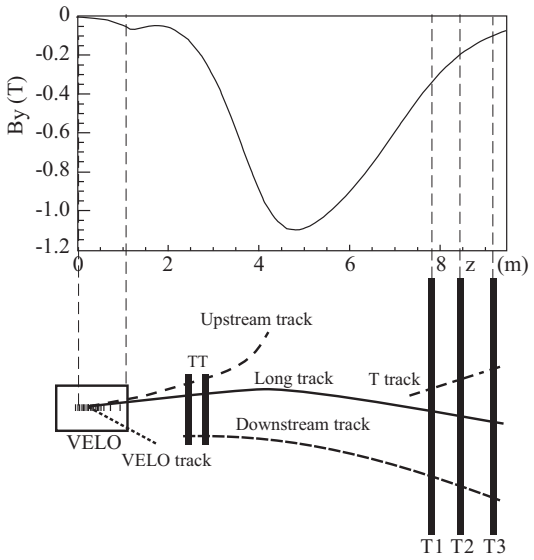

Figure 3.9: Magnetic field strength along the z axis (beam direction) of the LHCb detector [93]. Also shown are the locations of the tracking sub-detectors with respect to the field strength and the types of tracks reconstructed in LHCb.

the magnet close to their photon detectors.

To achieve the maximum momenta precision, the magnetic field of the magnet must be known to a high precision. The magnetic field along the z axis of the detector as measured using an array of Hall probes and is shown in Fig. 3.9. The momentum scale can be further corrected in the data by scaling the measured momentum of tracks to reproduce the known J/ψ mass [94]. This is done on a run by run basis. Over the course of data taking, the magnet polarity is regularly flipped such that approximately the same size data samples are collected for each magnet condition. This is done in an effort to reduce instrumental asymmetries in measurements of CP violating asymmetries where the difference between a final state and its charge conjugate is measured.

3.2.5 Combined tracking performance

Tracks in LHCb are treated in different categories depending on how many of the tracking sub-detectors they pass through. The lower part of Fig. 3.9 shows an illustration of all of these track types. The tracks with the most precisely measured momenta are long tracks, tracks which have hits in both the VELO and the tracking stations. They are the most common track type for most decays of interest of LHCb as the B meson is expected to decay within the VELO. Due to having the best momenta resolution, these tracks are the most useful for physics analyses. In

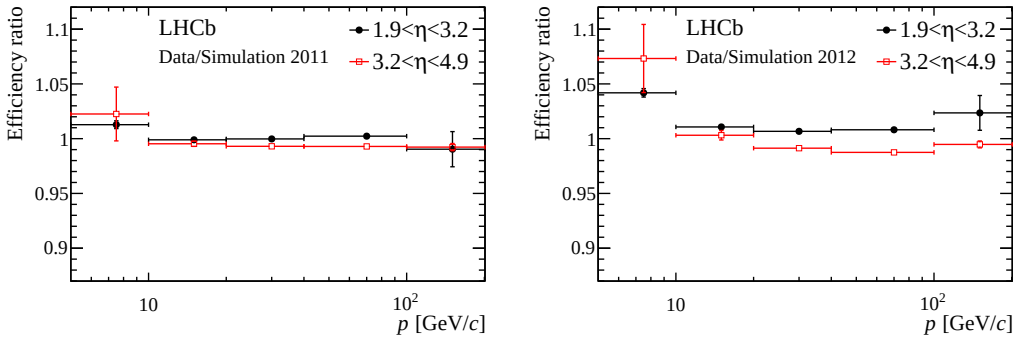

Figure 3.10: Ratios of tracking efficiency from data to simulation in bins of track momenta and pseudorapidity for 2011 (left) and 2012 (right) conditions [95].

addition there are VELO tracks which only have hits in the VELO, upstream tracks which are low momentum tracks with hits in the VELO and TT before being swept out of the detector in the magnet region, downstream tracks which have TT and tracking station hits and generally result from decays of long lived particles outside of the VELO, and T tracks which only have tracking station hits and tend to come from secondary interactions. All of the tracks described above need to be reconstructed from the hits in the tracking sub-detectors. Initially, seeds of a track are found by searching for straight line segments of hits in the individual sub-detectors where the magnetic field is low. These seeds are then matched together and a Kalman filter used to refit the complete track trajectories, accounting for additional affects such as multiple scattering of particles and energy loss in the detector.

The total tracking efficiency is calculated on the reconstructed tracks, using information from all sub-detectors. When carrying out analyses it is important that these efficiencies can be correctly modelled in simulation. To validate this, these efficiencies can be evaluated for selected tracks from data and compared to simulation [95]. In the same way as is done for the TT performance measurements described in Sec. 3.2.3, the decay $J/\psi \to \mu^+\mu^-$ is used to measure these efficiencies using a tag-and-probe technique. By taking one of the muon tracks as a fully reconstructed tag to select out the events and then only partially reconstructing the other muon track and attempting to match it to a fully reconstructed long track in the data, the reconstruction efficiency can be determined. The results of this, shown as the ratio of the efficiency taken from data divided by the efficiency taken from simulation in bins of track momenta and pseudorapidity are shown in Fig. 3.10. In most regions there is good agreement between data and simulation. Simulated samples are corrected for these discrepancies when performing analyses.

3.2.6 Particle identification

One of the key strengths of the LHCb detector is its ability to identify different types of charged particles. Distinguishing between kaons, pions and protons is vital to be able to carry out the majority of flavour physics measurement that the detector is designed for.

There are three sub-systems involved in particle identification (PID), the ring-imaging Cherenkov (RICH) detectors, calorimeters and muon detector systems that combined give PID information required to select pions, kaons, protons, electrons and muons that are produced in decays, as well as for photons and neutral pions.

3.2.7 RICH detectors

The primary PID information for charged hadrons is provided by the RICH detectors. These make use of the Cherenkov effect, where when a particle passes through a medium with a velocity greater than the speed within that medium, Cherenkov radiation is emitted as a cone of light in the forward direction with respect to the particle. The opening angle of this cone is dependent on the refractive index of the medium that the particle travelled through and the velocity of the particle. The general design of the RICH detectors is to make use of a radiator medium in which Cherenkov photons are produced, measuring the Cherenkov angle and using particle momenta measurements to determine the mass and thus the species of the particle. In practice, in LHCb, the Cherenkov angle is not ever directly determined. The likelihood for a particle to be a given hypothesis is instead evaluated by comparing the pattern of hits in the detector against the expected pattern of hits for each hypothesis. The produced Cherenkov photons are reflected and focussed by an optical system of mirrors and then detected by Hybrid Photon Detectors (HPDs). The performance of these HPDs are adversely affected by magnetic fields and, as mentioned in Sec. 3.2.4, the detectors are surrounded by an iron shield to reduce any magnetic field within the sub-detectors. The HPDs are also individually shielded using a material with high magnetic permeability. Schematic views of both RICH1 and RICH2 are shown in Fig. 3.11.

The RICH system comprise two different RICH detectors, each one optimised to operate in a different momentum range so as to provide good coverage of the momentum range $2 - 100\,\mathrm{GeV}/c$. The RICH1 sub-detector makes use of a C_4F_{10} radiator, covering the low range of the momenta spectrum from $2 - 40\,\mathrm{GeV}/c$. In addition, in Run 1 of the LHC, an aerogel radiator was also included which provided

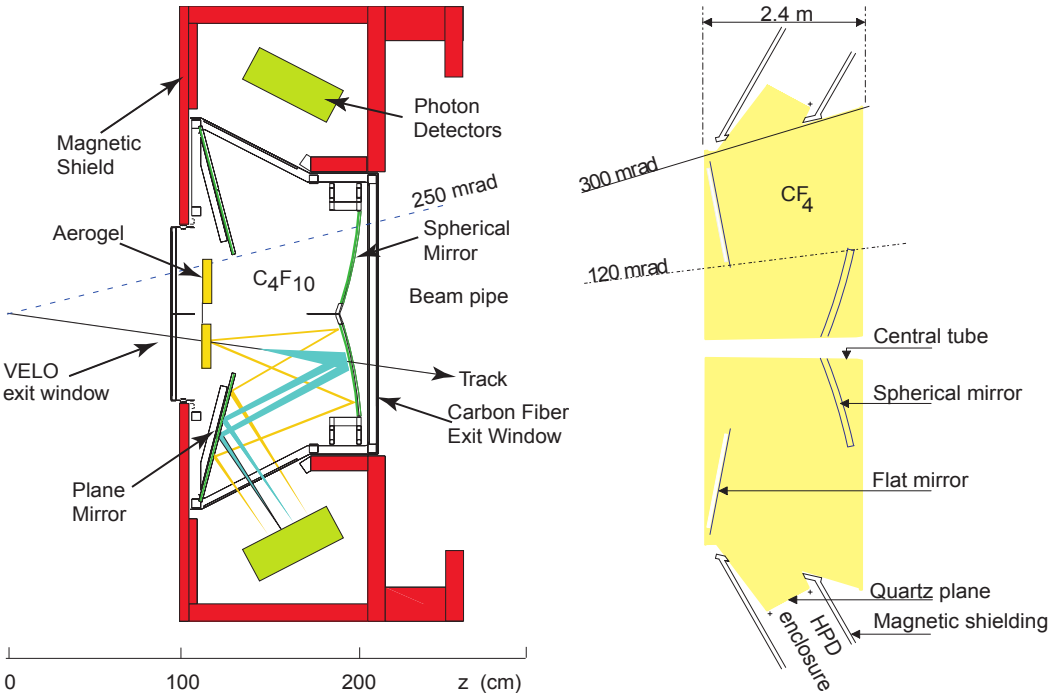

Figure 3.11: Schematic view of RICH1 (left) and RICH2 (right) [83]. RICH1 is viewed from the side whereas RICH2 is viewed from above.

coverage to even lower particle momentum. The sub-detector is located upstream of the detector magnet and covers the full acceptance of the detector. The high end of the momenta spectra from $15 - 100\,\mathrm{GeV}/c$ is covered by RICH2 which makes use of a CF_4 radiator. This is located downstream of the magnet and covers a limited acceptance range, however it is in this range that the high momentum particles are produced.

The Cherenkov angle over the momenta range of the experiment for just the RICH1 C_4F_{10} radiator material as measured in Run 1 data is shown in Fig. 3.12. Good separation is achieved between the different particle species hypotheses. Unfortunately, the performance of the aerogel did not match design expectations and the Cherenkov angle resolution is worse than that expected from simulation. This is believed to be due to the aerogel structure absorbing some of the RICH1 gas radiator and distorting the expected Cherenkov angle.

The aerogel radiator in RICH1 was removed for Run 2 [96], principally because the gain in particle identification performance turns out to be small in a high track multiplicity environment. By removing the aerogel, the speed of reconstruction in the RICH is greatly increased due to a reduction of the amount of photon candidates in the detector. In addition, the removal of the aerogel allows parts of the

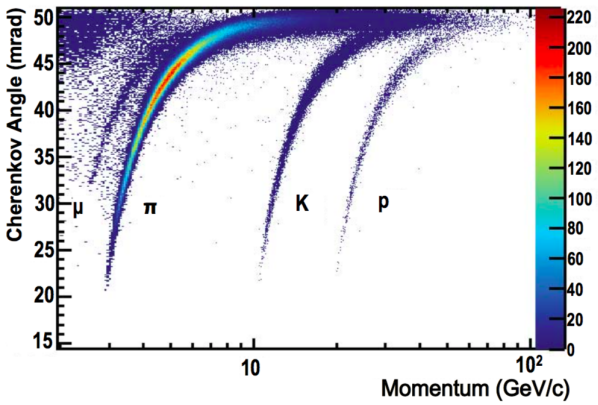

Figure 3.12: Cherenkov light angle as a function of track momentum in the RICH1, C_4F_{10} radiator as measured from Run 1 data [97]. Bands for various charged particles are clearly visible and labelled.

gas radiator that had previously been blocked by the aerogel to be fully exploited. A number of HPDs have also been exchanged for refurbished detectors which are of higher quality and with readout electronics more suitable for the LHCb environment. The combination of these changes has lead to an improvement of pion-kaon separation information from the RICH systems.

3.2.8 Calorimeters

The calorimeter system is made up of an electromagnetic calorimeter (ECAL) followed by a hadronic calorimeter (HCAL), with a preshower detector (PS) and scintillator pad detector (SPD) placed before these as part of the system. Energy deposition in each part of the calorimetry system is detected from scintillation light being transmitted to photomultipliers. The SPD is placed at the front of the system and is used to select charged particles. Charged particle will deposit energy in the the SPD while neutral particles, such as the γ and π^0 will pass through before depositing energy in the ECAL. The PS and SPD are separated by a thin layer of lead which acts as a converter, with electrons being caused to shower by the converter and being detected by the PS while charged hadrons pass through before depositing energy in the HCAL.

All of the calorimeters adopt a variable cell density across the active detector surface, so as to have maximum sensitivity in the region with the highest particle flux (closest to the beam line). The layout of a single quadrant section for both the ECAL and HCAL is shown in Fig. 3.13, with the cell size and number of readout channels given within this section. The SPD/PS share the same layout as the ECAL.

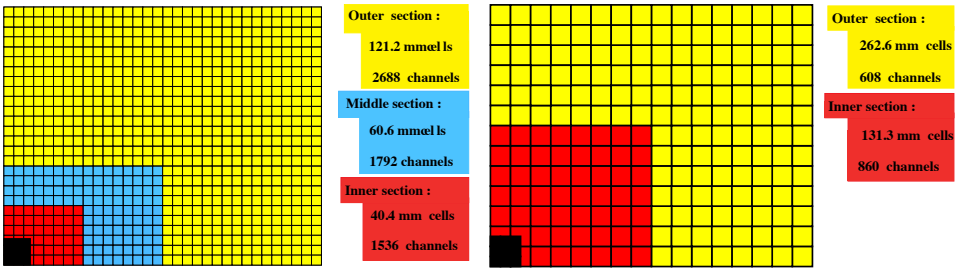

Figure 3.13: Representation of the segmentation for the ECAL (left) and HCAL (right), with the cell dimensions and number of channels given in each region [83].

The ECAL provides useful information for the electron identification and reconstruction, but its key importance is in measurements involving photons and neutral pions. No other part of the detector provides energy and momentum measurements of these particle types. The primary purpose of the SPD/PS part of the calorimeter is to provide fast particle identification information for the hardware trigger (discussed in detail in Sec. 3.2.10) of the experiment. Depending on the relative energy deposition in the SPD, PS, ECAL and HCAL and the structure of the resulting shower, a track can quickly be identified as a photon, electron or hadron.

3.2.9 Muon stations

A key component necessary for the use of the LHCb detector is an effective muon identification system. This is provided by a five station (M1 – M5) muon detector system which provides vital, fast early information for the hardware trigger as well as the muon identification used in the high level software triggers and for analysis work. The detectors are made up of multi-wire proportional chambers which are partitioned into logical pads that provide single point hit information in each detector station.

The five stations are arranged such that the first station, M1, is located upstream of the calorimeters and the other four stations are located downstream and are shown in Fig. 3.14. M1 is positioned in this way to improve the measurement of muon p_{T} that is used in the hardware trigger. The stations M2 through to M5 are placed sequentially without any other active detector components between them, however there are 80 cm thick iron absorbers placed between each station. These are used to select out penetrating muons as they pass through the detector. The inner acceptance of the muon system is 20 (16) mrad and the outer acceptance 306 (258) mrad in the x (y) direction of the detector respectively. A constant angular acceptance necessitates that the size of the stations in the x-y planes increases the

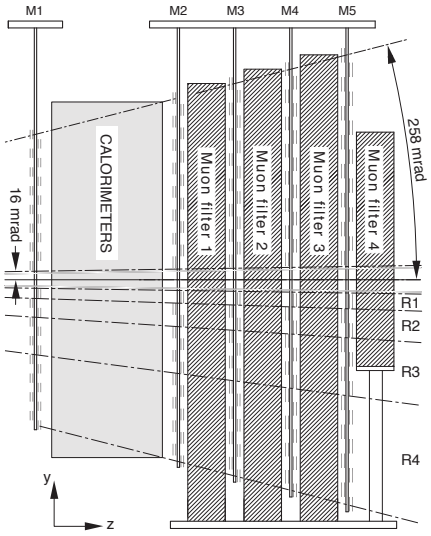

Figure 3.14: Schematic view of the muon system from the side [83]. Shown are the stations, M1-M5, the location of the calorimeters and iron filters between the latter stations.

further downstream in the detector they are located.

Muon stations M1 – M3 have a higher spatial resolution then the later station, allowing the p_{T} of muon tracks to be measured with an approximately 20% resolution using just the muon detectors, where this resolution is defined by the dimensions of the logical pads in the detector chambers. The muon systems on their own are used in the hardware trigger stage to select out events containing muons. For this trigger requirement to be fulfilled, five hits across the stations are needed. In addition, the muon systems feed into the global particle identification algorithms alongside the other PID detectors as well as providing information for final analyses on the final muon momenta.

The LHCb experiment aims for a 95% efficiency of the hardware muon trigger, thus requiring that each individual muon station has an average efficiency of approximately 99%. This efficiency has been measured in data taken in 2010, 2011 and 2012 [98]. The efficiency is measured for each of the muon stations, divided by region within each station, and is shown in Fig. 3.15. It is clear that for the majority of regions the hit efficiency exceeds 99% and when taking into account the total response of the muon system the trigger efficiency does exceed the aimed for 95%.

Combined particle identification

For the majority of use in offline analyses, PID information taking from all of the individual sub-detectors is combined to produce simple variables that represent the

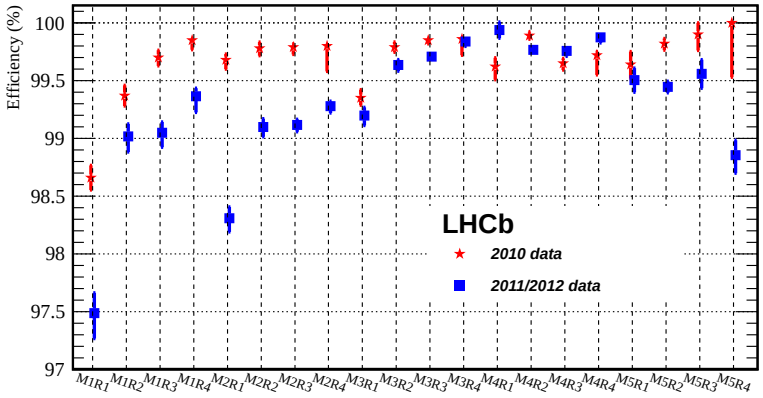

Figure 3.15: Measured hit efficiency for different stations and regions of the muon system, measured for both 2010 and 2011+2012 combined [98].

likelihood that a given track is one of a proton, kaon, pion, muon or electron. This combination is generally carried out in one of two ways, either a linear combination of likelihoods from each sub-system or a more advanced combination using multivariate analysis techniques [93]. The linear combination gives a combined likelihood for a track being a proton, kaon, muon or electron relative to a pion hypothesis. The multivariate technique includes extra information from the whole event and accounts for correlations between information from the sub-detectors. The output of this technique is a single probability for each of the potential particle species that a track could be.

The performance of the multivariate variables should be an improvement over the combined likelihoods and the extent of this improvement can be measured in data. In Fig. 3.16 the performance of both the types of PID combination for both muon and proton identification are shown for a specific decay type. In this figure the area to the left of the data points can be interpreted as a measure of the performance of the variables, with a larger area indicating higher performance. It is clear that the use of the multivariate technique based variables provide a significant improvement in PID performance.

3.2.10 Trigger

The trigger system acts to reduce the data flow out of the detector to a rate that can be written to data storage and subsequently used for analysis purpose. The beam crossing rate of the LHC is 40 MHz however the LHCb detector can only be read out at a rate of 1 MHz and data is only saved to disk at a rate of between $2 - 5$ kHz in Run 1 and up to 12 kHz in Run 2. This reduction is done in a three step process via

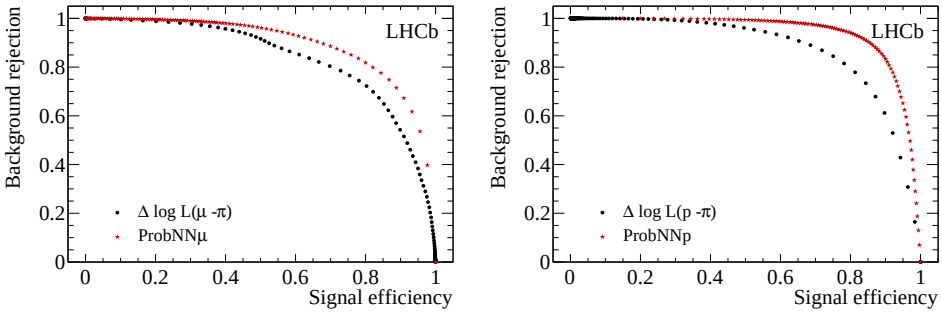

Figure 3.16: False identification rate verses correct identification efficiency for muons (left) and protons (right) from simulated $\Sigma^+ \to p\mu^+\mu^-$ decays [93]. Linear likelihood combination variables are shown in black, multivariate analysis techniques variables in red.

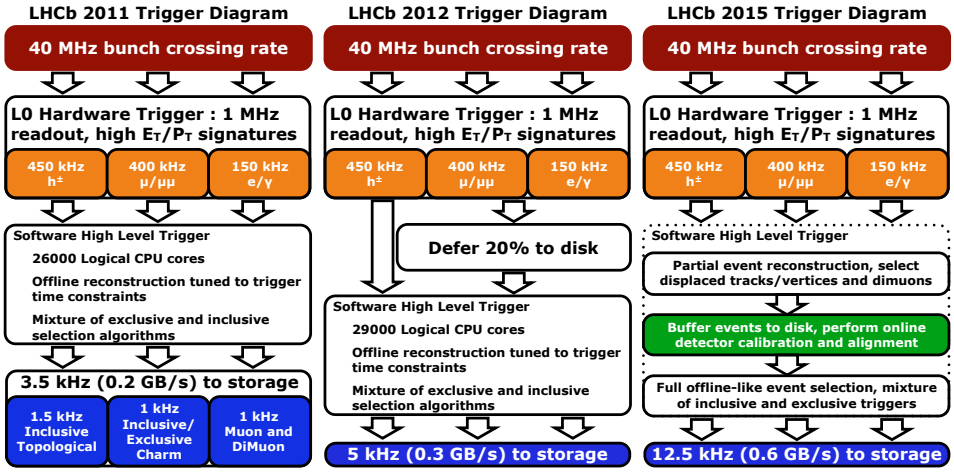

Figure 3.17: Trigger schemes for Run 1 - 2011 condition (left), Run 1 - 2012 conditions (middle) and Run 2 (right) [99].

a hardware based trigger (L0), first high level (HLT1) and second high level (HLT2) triggers. The general flow of information through the trigger system for 2011, 2012 and 2015 (where 2015 is representative of all Run 2) is shown in Fig. 3.17. Between Run 1 and Run 2, the flow of information through the trigger system has been changed to allow a more flexible approach to the use of the software trigger. Whereas in Run 1 the HLT1 and HLT2 software triggers ran constantly throughout data taking, in Run 2 a buffer has been introduced, allowing more efficient use of the experiments computing resources. By deferring the trigger decision the computing resources can be used at all times and not just during data taking. It is also possible to use this buffer to run data calibration tasks such that the most precise tracking/particle identification is available in HLT2.

Hardware trigger

The hardware part of the trigger (L0) relies entirely on decisions from the detector components to reduce the rate from the maximum bunch crossing rate of 40 MHz to 1 MHz, the upper limit at which data can be read out from the current detector. The L0 trigger decision is motivated by the fact that the final state particles from a B meson decay are likely to have high transverse energies (E_{T}) and momenta (p_{T}). As such, information from the calorimetry systems and muon systems are used to select these final state particles as identifiers of B decays.

For selection based on the E_{T} of particles, the calorimeters are used. Clusters in either the ECAL or HCAL with a total energy greater than the trigger requirement pass the L0 trigger. This trigger requirement is different for different particle types. As discussed in Sec. 3.2.8, the inclusion of the SPD/PS system is a vital inclusion to be able to distinguish electron, photons and hadrons and the hardware trigger level. Muons are selected to pass the L0 trigger based on the p_{T} of the particles, with this decision being made purely using the muon system and a momentum kick approximation for the magnet. The two highest p_{T} muons are selected and these p_{T} values are then compared to a threshold value which makes the decision either based on a single muon, the `L0Muon` trigger, or the product of the two muons, the `L0DiMuon` trigger. Typical values of these thresholds in Run 1 are $p_{\mathrm{T}} > 1.76\,\mathrm{GeV}/c$ for the single muon trigger and $p_{\mathrm{T}1} \times p_{\mathrm{T}2} > (1.60\,\mathrm{GeV}/c)^2$ for the dimuon trigger. The p_{T} threshold to trigger can be varied and the characterisation of these changing thresholds are an important consideration in analyses. The introduction of a buffer in Run 2 has lead to more frequent updates of the L0 thresholds over data-taking periods to make sure data can be adequately processed without filling the available disk space.

Software trigger

After passing the hardware trigger, events are passed to the software stage known as the High Level Trigger (HLT). This consists of two stages, HLT1 and HLT2, which is carried out by a C++ applications running on an approximately 30000 CPU core computing farm. A decision on whether to keep an event is made in multiple stages, at each point only using a subset of the available information in order to make the most efficient use of the computing resources.

The initial purpose of HLT1 is to perform a simple track reconstruction of the events triggered at L0 and confirm the L0 decision. The VELO pattern reconstruction software is fast enough to be run on each event passed to HLT1,

allowing a reconstruction of VELO tracks and thus vertices to be made. Tracks are then selected based on their impact parameter with respect to any primary vertex, the χ^2 of a linear fit to potential track (in the case where there are muon hits) and the momenta and transverse momenta of the tracks.

After passing from HLT1 to HLT2, the rate of events has been reduced to such a level that a full event reconstruction can be carried out. Forward tracking of all VELO tracks, where a VELO track is matched with hits in the detectors further downstream in the detector, is carried out. Only tracks over a certain momenta and transverse momenta, typically around $3\,\mathrm{GeV}/c$ and $0.3\,\mathrm{GeV}/c$ respectively, are reconstructed to increase the speed of the algorithm. At this point many different trigger selections are carried out which are motivated by the requirements for different physics analyses. The output from these final selections are then stored to disk for use in offline analysis.

In Run 2, all events that pass the HLT1 trigger selection are buffered rather then immediately being passed to the rest of the software trigger and data storage. This allows for less strict software trigger requirements as the decision of HLT can be deferred as long as there is still space in the buffer. This also has the advantage of allowing the software trigger selection to be carried out using tracks/PID information with the same precision as the offline selection. The design of this system makes use of LHC downtime where more of the computing nodes are available to run the trigger software and can thus process the excess events in the buffer before the next period of data taking.

3.3 Future upgrades of LHCb

While the performance of the LHCb detector has been good in the first two runs of the LHC, there are currently limitations on the quality and quantity of the data that can be taken with the detector. In particular, the necessity of a tight hardware trigger due to the limitation of data readout speed from the detector provides a hard limit on the total size of datasets that can be recorded for any given time period. Between Run 2 and Run 3 of the LHC, the LHCb detector will undergo an upgrade with the intention to reduce these limitations [100, 101].

The key feature of this upgrade will be the complete removal of the hardware trigger from the experiment, transitioning to a fully software based trigger. This requires the experiment to read out at the full 40 MHz resulting from the beam crossing frequency. In addition, all key sub-detectors will undergo individual upgrades to various extents. These upgrades are motivated by the need to upgrade the readout electronics of each sub-detector, changes to cope with the different running environment expected in Run 3 and by the need to replace existing detector components that will have undergone degradation due to radiation damage. In terms of running conditions, LHCb will run with an operational luminosity of $\mathcal{L} = 2 \times 10^{33}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ that will increase the occupancy of the sub-detectors, requiring new detectors in areas that are already at occupancy limits. A representation of the key changes to the detector design is shown in Fig. 3.18.

The current VELO will no longer be suitable for LHC runs after Run 2 due to degradation from radiation damage. The VELO will be upgraded, with a move from the current silicon strip based detector to one based on silicon pixels [103]. These hybrid pixel sensors will be paired with new read-out electronics with a radiation hardness that will be able to withstand the Run 3 luminosities. The cooling system of the new VELO will be integrated into the VELO modules and will be an upgrade over the current cooling, designed to deal with the heat from both the new module mounted read-out electronics as well as significant radiation in the active region of the sub-detector. While the general design of the pixel sensors that will be used has been decided on, multiple designs for the support structure for each upgrade VELO module and on the cooling system that will be used have been proposed. To aid in the decision making process on which of these designs to use, simulation generated using models of the upgrade VELO with these different proposals have been carried out. Appendix A gives the details of the geometry description of potential designs for the upgraded VELO and how these have been used to contribute to the design decisions.

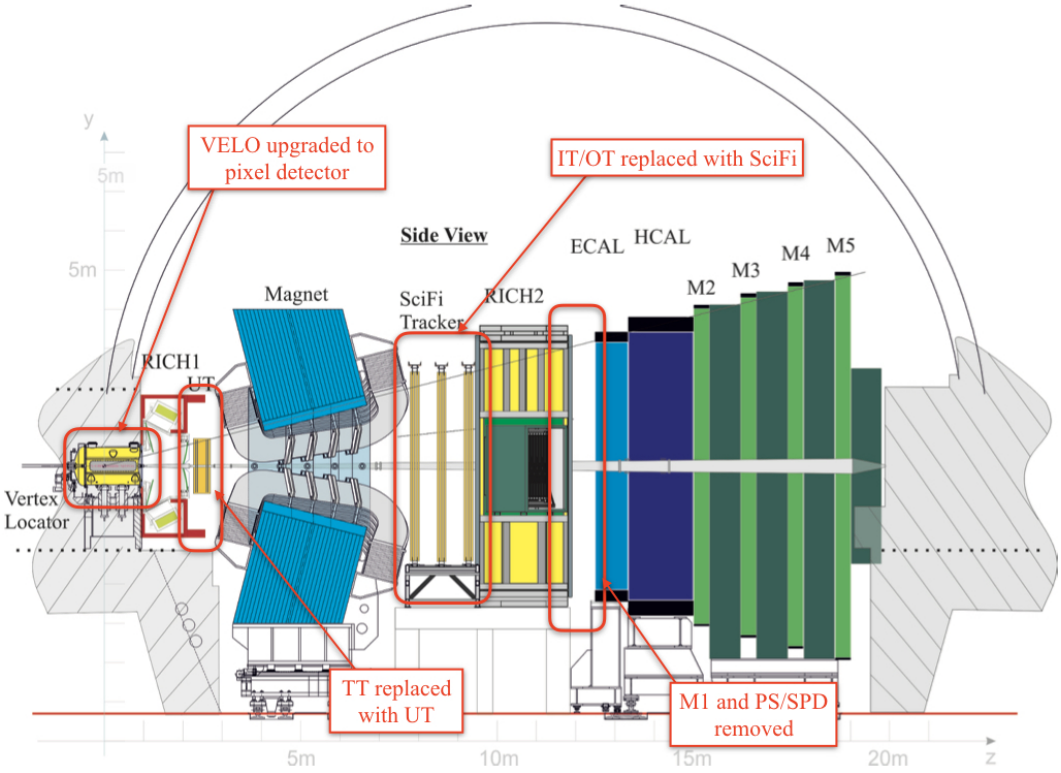

Figure 3.18: A cross section of the LHCb detector upgrade design, as viewed from the side. Labelled are the key changes with respect to the current detector. Image adapted from [102].

For the other sub-detectors in the tracking system, the TT will be replaced with the UT and the IT/OT (T1 – T3) systems will be replaced with the scintillating fibre (SciFi) tracker [102]. The TT sensors were not designed with the radiation environment for Run 3 in mind and the sensors close to the beam are not sufficiently radiation hard. The UT makes use of the same technology as the TT, using planes of silicon strips to provide tracking measurements. It is an improvement over the TT as it makes use of thinner silicon sensors, reducing the material in the detector and improving measurement precision. The segmentation of the sensors is finer, giving a better position resolution, and the sensors in the inner region have been designed to withstand the radiation environment up to integrated luminosities of $50\,\mathrm{fb}^{-1}$. In addition, the sensors are designed to placed closer to the beam pipe, providing better acceptance for tracks at high pseudo-rapidity than the current TT.

The current tracking stations are limited by the maximum operational occupancy of the OT stations. In the current detector, the area covered by the IT is chosen so as to limit the occupancy of the OT, however this is only valid for the luminosities used in Run 1 and Run 2. The increased Run 3 luminosity would lead

to OT occupancies which would severely impact track finding efficiency. As some of the OT would need to be replaced and all of the read-out electronics would need to be replaced for both IT and OT to enable 40 MHz read-out, the decision has been made to fully replace the TT stations with a new tracking system. The IT and OT will be replaced with a scintillating fibre tracker with fibres with a diameter of 250 µm, arranged in the same layered structure as the current silicon tracker. Silicon photomultipliers will be placed at the ends of the scintillating fibre modules to read-out the detector.

For the most part, the PID systems will remain similar to the current detector design [104]. The main upgrade of the RICH will be transition from HPD to multi-anode photomultipliers (MaPMTs) which will have suitable readout electronics for the 40 MHz conditions, with the remainder of the RICH detectors remaining the same. In the calorimeters, the main change will be the removal of the PS and SPD detectors from in front of the ECAL. As discussed in Section 3.2.8, the main purpose of these detectors is to provide information for the L0 trigger, which will not be used in the upgrade. In addition, station M1 of the muon system will be removed as with the upgraded trigger with no L0 trigger, an initial p_{T} measurement is no longer needed.

With the move to a fully software based trigger, the computing resources and software architecture will need to be reevaluated and provide a great challenge for the detector upgrade [105, 106]. By reading out all events and using just a software trigger, the processing speed of the software must be vastly improved from the current 1 MHz to the required 30 MHz. Note the difference between the read-out rate of 40 MHz and the required rate of the trigger of 30 MHz. The read-out rate corresponds to the maximum data rate defined by the bunch crossing frequency, whereas the trigger rate is reduced due to the gaps in the bunch filling scheme used by the LHC. The rate at which data is read-out and stored from the trigger must also be upgraded, with a total event output rate from the trigger to be stored of 50 kHz required to maintain a similar broad physics programme as has been carried out in Run 1 and Run 2. In the upgrade trigger, HLT1 will act in much the same way as in Run 2, reducing the event rate by selecting signal like events using a fast track reconstruction and applying selection criteria. This is expected to reduce the rate from 30 MHz to around 1 MHz which will be buffered and then processed by HLT2. A full reconstruction will then be carried out on the events, with selection criteria applied based on specific trigger lines which correspond to specific physics analyses. This allows only the selected candidates to be stored while removing the rest of the event, decreasing the data storage requirements while streaming the triggered events

in datasets that can then be accessed by analysts. Development of the trigger to run at increased rates and ensuring that the selection criteria used in the trigger are properly validated before the start of data taking are vital tasks during the upgrade period.

More long term upgrades are also being considered for the LHCb experiment [107,108], with the main focus being on how to take advantage of the increased luminosities available once the LHC is upgraded to the High Luminosity LHC.

Chapter 4

Search for the decay $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$

The following chapters present the results of a search for the rare decay $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$. In particular, in this chapter an introduction and background to the decay mode is given and the analysis strategy to search for the decay mode and measure its branching fraction is presented.

4.1 Introduction

As previously discussed in Sec. 2.3.6, rare $b \to d\ell^+\ell^-$ transitions are of great current and future interest in the search for NP effects. The large LHC datasets have made it possible to observe decays involving this type of transition. Only two decay modes of this type have previously been observed, with the LHCb collaboration observing both the decays $B^+ \to \pi^+\mu^+\mu^-$ [109] and $\Lambda_b^0 \to p\pi^-\mu^+\mu^-$ [110]. For the decay $\Lambda_b^0 \to p\pi^-\mu^+\mu^-$, the dataset comprises a mixture of resonances in the $p\pi$ system that are experimentally hard to separate and thus challenging to compare against theoretical predictions. The LHCb collaboration has also provided evidence for the decay mode $B^0 \to \pi^+\pi^-\mu^+\mu^-$ [111], where the decay $B^0 \to \rho^0\mu^+\mu^-$ is expected to dominate in the $\pi^+\pi^-$ invariant mass range of 0.5 – $1.3\,\text{GeV}/c^2$ the measurement is performed over. Expanding the range of modes that have been observed and, when large enough data samples have been collected, carrying out more detailed angular analyses similar to those performed for $b \to s\ell^+\ell^-$ transitions is very important in the field of rare flavour physics.

The decay $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ is particularly of interest due to its relation to the decay $B^0 \to K^{*0}\mu^+\mu^-$, being the $b \to d\ell^+\ell^-$ partner to this decay. The leading order

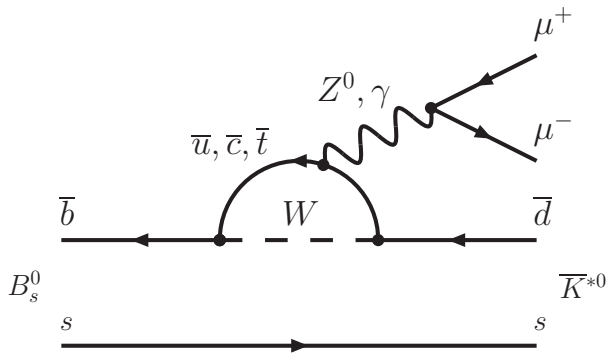

Figure 4.1: Leading order Feynman diagram for the decay $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$.

Feynman diagram for this decay is shown in Fig. 4.1. As previously discussed in Sec. 2.3.7, evidence of anomalies in angular observables of the $B^0 \to K^{*0}\mu^+\mu^-$ decay motivates investigation of the more suppressed mode, both due to the potential for larger effects and so as to investigate the flavour structure of any potential NP effect.

Predictions of the branching fraction of this decay mode according to the Standard Model are in the approximate range of 3–4 $\times 10^{-8}$ [112–114], where the range of predictions results from the variety of possible techniques and assumptions involved in form factor calculations as discussed in Sec. 2.2.2. This leads to the expectation, taking into account the ratio of B_s^0 to B^0 production at the LHC, that this should occur at the approximate rate of 1 decay for every 100 $B^0 \to K^{*0}\mu^+\mu^-$ decays. The heavy suppression of this decay mode makes it a particularly challenging mode to investigate. This is complicated by the fact that the only way to separate the decay from $B^0 \to K^{*0}\mu^+\mu^-$ is by the mass of the B_s^0 and there is only a small mass difference (87.5 MeV/c^2). While the GPDs at the LHC have been able to make measurements of the $B^0 \to K^{*0}\mu^+\mu^-$ mode, only the LHCb detector has the invariant mass resolution capable to separate the two modes.

In this thesis, the first search for the $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ decay mode is presented making use of a combined dataset of 1 fb^{-1} of data collected by LHCb in 2011 at a c.o.m. energy of 7 TeV, 2 fb^{-1} collected in 2012 at a c.o.m. energy of 8 TeV and 1.6 fb^{-1} collected in 2016 at a c.o.m. energy of 13 TeV. Henceforth, the combined 2011 and 2012 dataset will be referred to as the Run 1 dataset while the 2016 dataset will be referred to as the Run 2 dataset. Whereas the dataset referred to as "Run 1" is the full LHCb Run 1 dataset collected, the "Run 2" dataset is only a subset of the full LHCb Run 2 dataset that will be available for analysis in the future.

4.2 Measurement strategy

The search for the decay mode $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ is carried out by performing a maximum likelihood fit to the $K^+\pi^-\mu^+\mu^-$ invariant mass of the reconstructed candidates. The significance of any observed excess of $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ events with respect to the null hypothesis is then evaluated via Wilks' theorem [115], using the difference in the log likelihood between the null hypothesis and the standard fit. In the event that a significance is greater than 3, it is considered to be evidence for the decay mode (and greater than 5 is considered to be an observation) and a measurement of the branching fraction of the decay mode is made. Otherwise, a limit of the branching fraction of the decay mode would be set. This analysis was carried out as a blind analysis, with the selection criteria and fit strategy chosen and optimised without examining the $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ signal region until the final measurement is made. This was carried out to ensure minimal bias in the measurement strategy.

The parameter of interest in the mass fit is the yield of $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ candidates and this can be converted to a measurement of the branching fraction by carrying out the fit with respect to a control mode with a measured branching fraction. The control mode used in this analysis is the decay $B^0 \to J/\psi K^{*0}$ with $J/\psi \to \mu^+\mu^-$, chosen due to having the same four body final state as the searched for mode and due to it having a well measured branching fraction to measure the $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ branching fraction relative to. The similarity of the final state particles between the modes is desirable as it allows most systematic uncertainties related to the two decay modes to cancel in a ratio between the two. The branching fraction for the decay $B^0 \to J/\psi K^{*0}$ is taken as

$$\mathcal{B}(B^0 \to J/\psi K^{*0}) = (1.19 \pm 0.01 \pm 0.08) \times 10^{-3} \quad ,$$

from Ref. [116]. This branching fraction is then multiplied by

$$\mathcal{B}(J/\psi \to \mu^+\mu^-) = (5.961 \pm 0.033) \times 10^{-2} \quad ,$$

from Ref. [2].

To perform this normalisation, a simultaneous fit to the control mode needs to be performed while production rates of the B^0 and B_s^0 mesons as well as the relative efficiencies of the two modes to pass selection criteria are taken into account. The correction for the B_s^0 and B^0 meson relative production ratio, f_s/f_d, is taken from

the LHCb measurement [117]

$$\frac{f_s}{f_d} = 0.259 \pm 0.015$$

and the evaluation of the efficiency ratio between the modes is detailed Chapter 6.

The full expression used in the branching fraction measurement is given by:

$$\frac{\mathcal{B}(B_s^0 \to \overline{K}^{*0}\mu^+\mu^-)}{\mathcal{B}(B^0 \to J/\psi K^{*0})\mathcal{B}(J/\psi \to \mu^+\mu^-)} = \frac{N(B_s^0 \to \overline{K}^{*0}\mu^+\mu^-)}{N(B^0 \to J/\psi K^{*0})} \times \frac{\epsilon^{\text{total}}_{B^0 \to J/\psi K^{*0}}}{\epsilon^{\text{total}}_{B_s^0 \to \overline{K}^{*0}\mu^+\mu^-}} \times \frac{f_d}{f_s}\,. \tag{4.1}$$

In addition to a measurement of the branching fraction determined as detailed above, two additional branching fraction ratios are considered: the ratio of $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ to $B^0 \to K^{*0}\mu^+\mu^-$ and the ratio of $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ to $B_s^0 \to J/\psi\overline{K}^{*0}$. The first of these is measured as the initial step towards determining the ratio of CKM factor V_{td}/V_{ts}, one of the differing factors in the Feynman diagrams describing these decays along with the hadronic form factors. A full determination of this ratio is not carried out as this requires more detailed input from the theory community to correctly describe the contributions associated to hadronic form-factors of the spectator quarks in the decay. The second of these is determined as an alternative measurement of the branching fraction with a slightly different cancellation of systematic uncertainties in the branching ratio due to being between two B_s^0 modes, at the cost of an increased statistical uncertainty resulting from the relatively smaller control mode yield.

Chapter 5

Selection and backgrounds

As detailed in the previous chapter, the decay $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ is predicted to be very rare in the SM with a branching fraction $\mathcal{O}(10^{-8})$. An efficient selection and careful consideration of potential backgrounds is essential to maximise the sensitivity in the search for the decay. This chapter provides details on the selection of candidate events and the backgrounds considered in the analysis.

5.1 Selection

In performing a search for a heavily suppressed decay it is absolutely vital that a well considered and efficient selection is carried out on the candidates. Of particular importance is the suppression of backgrounds with minimal impact on the signal. Here, a multi stage selection process is detailed and the key features of each stage of the process and how this provides the best selection for this search is explained. A schematic illustration of the LHCb data flow and selection process is shown in Fig. 5.1.

5.1.1 Trigger

As detailed in Sec. 3.2.10, LHCb makes use of a combined hardware and software trigger to reduce data flow and select out relevant events. In addition, as the trigger has been updated between Run 1 and Run 2 [99] the requirements on the trigger are slightly different between the two datasets, although applied for the same reasons. The trigger requirements as named within the LHCb software for both datasets are given in Tab. 5.1.

For Run 1, only the `L0Muon` trigger requirement is applied, a requirement that a single high p_T muon is triggered on in the event. The `L0DiMuon` requirement,

triggering on the product of p_T of each muon in a two muon event, is not applied here due to it adding very little to the dataset. This is, however, not true for the Run 2 dataset where a higher p_T threshold was required on the single muon trigger. In the L0 trigger, a selection is also made to reject busy events where the detector occupancy is high, making track reconstruction more difficult and taking more time in the high level trigger. This is referred to as the global event cut (GEC) and the SPD detector hit multiplicity is used as an estimate of the number of charged particles in an event. The inclusion of the extra `L0DiMuon` line allows access to additional events, approximately 24% more, and therefore the trigger line is included. There is, however, a difference in the GEC applied as part of each trigger line which is tighter for the `L0Muon` line then for the `L0DiMuon` line in Run 2, where the requirements are < 450 and < 600 respectively. This difference needs to be considered when carrying out an analysis.

The HLT1 trigger lines select out events as described in Sec. 3.2.10, selecting events containing one or more tracks with a high IP or p_T. A looser set of p_T and IP requirements are applied if the track matches hits in the muon system. The key trigger requirement at the HLT2 level is a series of topological triggers which select candidates that are topologically similar to a b-hadron decay; that is the candidate has a displaced secondary vertex and final state tracks with high IP and p_T. As in HLT1, if one of the daughters is identified as a muon a looser set of selection criteria are applied. The final requirement is the dimuon detached trigger, this makes use of information on the decay length of the dimuon to select low mass dimuon candidates. Note that all candidates used in this analysis are required to have at least one track that has triggered one of the trigger decisions at each of the L0, HLT1 and HLT2 stages detailed in Tab. 5.1. This is referred to as trigger on signal (TOS) rather

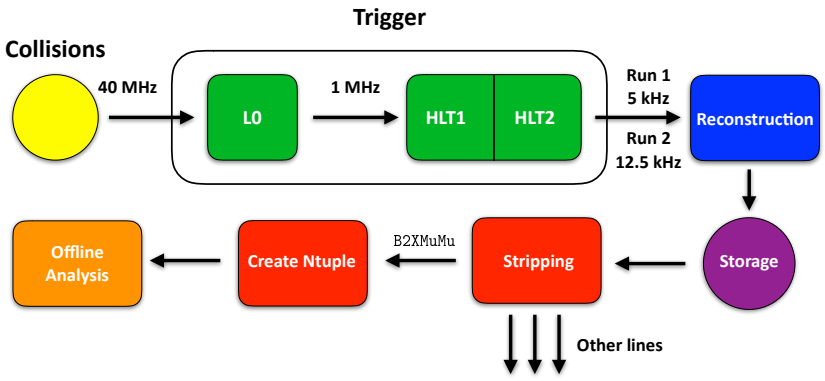

Figure 5.1: Schematic representation of the LHCb data flow and selection process used in this analysis.

than trigger independent of signal (TIS), where candidates could be selected by any other track in the event being triggered on.

Table 5.1: Trigger requirements applied to the Run 1 and Run 2 datasets. At all stages the trigger requirements are required to be passed by the decay products of the signal decay.

Level	Requirement	
	Run 1	Run 2
L0	L0Muon	L0Muon or L0DiMuon
HLT1	Hlt1TrackMuon or Hlt1TrackAllL0	Hlt1TrackMVA or Hlt1TwoTrackMVA or Hlt1TrackMuon
HLT2	Hlt2Topo{2,3,4}BodyBBDT or Hlt2TopoMu{2,3,4}BodyBBDT or Hlt2DiMuonDetached	Hlt2Topo{2,3,4}Body or Hlt2TopoMu{2,3,4}Body or Hlt2TopoMuMu{2,3,4}Body or Hlt2DiMuonDetached

5.1.2 Stripping

To handle the extremely large datasets collected by the LHCb detector, a process known as "stripping" is used. Any individual analysis makes use of a single or series of dedicated stripping lines that apply loose cuts to select out events of interest while maintaining a high signal efficiency. These lines are collected into streams of similar decay types and run centrally by the collaboration over the full dataset collected (on a year by year of data taking basis). For example, the candidates in this analysis were reconstructed in a dilepton stream alongside a number of different decays with dilepton final states. In this way, more manageable datasets containing only events of interest to specific analyses are created, removing the need for each individual analysis to run over the entire LHCb dataset.

Before the stripping procedure can be carried out, a full reconstruction of the candidates selected from the trigger lines is performed. As detailed in Sec. 3.2.10, the HLT carries out a reconstruction of all the tracks in an event to apply selection criteria in the trigger lines. An additional reconstruction is then performed offline, taking into account calibration and alignment of the individual sub-detectors performed for the relevant period of data taking. The information from hits in individual sub-detectors is included to provide the position, energy and particle identification details for each individual track. A kinematic fit to the resulting candidates is then performed using a Kalman fit, simultaneously reconstructing the vertices of the decay chain from tracks of the candidates. This offline re-fitting procedure is carried out using the decay tree fitter algorithm [118]. In this analysis,

the candidate is constrained to have originated from its most likely primary vertex, the PV with which the reconstructed B meson has the smallest IP. In addition, a reconstruction under the hypothesis that the two muons in the system have decayed from a J/ψ resonance is carried out, constraining the invariant mass of the dimuon system to the known J/ψ mass.

For this analysis the stripping line used is the `B2XMuMu` line. This is designed to select out events where a b-hadron decays to oppositely charged muons plus additional particles by applying a set of common cuts as well as specific cuts based on the specific additional final state particle. The values of the cuts applied in the `B2XMuMu` line, where there is a K^{*0} reconstructed as a kaon and a pion, are given in Tab. 5.2. The simplest requirements applied on the variables in the stripping are on the reconstructed invariant masses of the B meson, the dimuon pair and the K^{*0}. Particle identification requirements are applied on the two muons, requiring them to have been identified as a muon by the muon system, the `IsMuon` criteria, as well as the likelihood that each track is a muon with respect to a pion hypothesis, $\mathrm{DLL}_{\mu\pi}$. Requirements are also made on the ghost probability of each track, an indication of the likelihood that any track is fake. Additional requirements are made on the quality of each vertex fit in the reconstruction, vertex χ^2/ndf, the distance between a decay vertex and its origin vertex divided by the uncertainty on the distance, flight distance χ^2 and the angle between the momentum of a reconstructed particle and the vector connecting the decay vertex to its origin vertex, the DIRA angle. Finally, requirements on the difference in vertex fit χ^2 when a track is added to the primary vertex, χ^2_{IP}, are made for the individual final state tracks. A requirement on the χ^2_{IP} of the reconstructed B meson is also applied. The purpose of the above selection criteria is to select out candidates where the B is significantly displaced from its best PV, is consistent with coming from that PV and where the tracks are consistent with having coming from a single 4-track vertex.

In addition to the cuts detailed as part of the stripping line, a number of additional cuts are applied during the construction of the dataset to obtain a cleaner data sample. A tighter window around the known K^{*0} mass is required then that in the stripping line, corresponding to a requirement that $826 < m(K^+\pi^-) < 966$ MeV/c^2. In addition, a requirement on the χ^2 of the result from decay tree fitter is made to ensure only events where the fit has been successful are included. This is important to limit the effect of badly reconstructed decays in the mass distribution of the signal. Finally, at this point it is useful to define two regions of dimuon invariant mass squared, q^2, in the data set. Candidates in the range $0.1 < q^2 < 19.0$ GeV$^2/c^4$ and excluding the regions $8.0 < q^2 < 11.0$ GeV$^2/c^4$ and $12.5 < q^2 < 15.0$ GeV$^2/c^4$, domi-

nated by the J/ψ and $\psi(2S)$ resonances respectively, are considered to have dimuon pairs which have come from non-resonant rare decays and henceforth are referred to as being in the rare mode. Candidates in the range $8.0 < q^2 < 11.0$ GeV$^2/c^4$ have dimuon pairs which have predominantly decayed from the J/ψ resonance and are used as a control channel, this region is referred to as the control mode.

Table 5.2: Stripping line selection.

Particle	Parameter	Cut Value
B	χ^2_{IP} (best PV)	< 16
	$m^{\text{reco}}_{B^0}$	$> 4600\,\text{MeV}/c^2$
	$m^{\text{reco}}_{B^0}$	$< 7000\,\text{MeV}/c^2$
	DIRA angle	$< 14\,\text{mrad}$
	flight distance χ^2	> 121
	vertex χ^2/ndf	< 8
$\mu^+\mu^-$	$m^{\text{reco}}_{\mu^+\mu^-}$	$< 7100\,\text{MeV}/c^2$
	vertex χ^2/ndf	< 9
K^{*0}	$m^{\text{reco}}_{K^+\pi^-}$	$< 6200\,\text{MeV}/c^2$
	vertex χ^2/ndf	< 9
	flight distance χ^2	> 9
tracks	ghost Prob	< 0.4
	min χ^2_{IP}	> 9
μ	`IsMuon`	true
	$\text{DLL}_{\mu\pi}$	> -3
GEC	SPD multiplicity	< 600

5.1.3 Simulated samples

In the following steps of the selection, it is a necessary to make use of a pure signal sample to optimise any applied criteria. This is provided through the use of Monte-Carlo (MC) simulated samples of the relevant decay modes produced with a series of software packages, the full chain of which is detailed in Ref. [119]. The start of the decay chain, the pp collisions, are generated with an LHCb specific version of Pythia [120, 121]. The actual decay of the hadronic particles are then modelled mostly with the EvtGen [122] package except for radiation from the final state particles which is handled by the Photos [123] package. Finally, the interaction with the detector material and a simulation of the data collection process is described using the Geant4 [124] package. In this analysis, simulated samples

of $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$, $B^0 \to K^{*0}\mu^+\mu^-$, $B^0 \to J/\psi K^{*0}$ and $B_s^0 \to J/\psi\overline{K}^{*0}$ events are used in the selection procedure and in the evaluating of efficiencies needed for the final calculation of a branching fraction (discussed in more detail in Chapter 6).

To validate how well the simulated samples model the signal, it is necessary to extract a pure signal sample from the data. This is done using the *sPlot* [125] method to extract *sWeights*. A fit to the $B^0 \to J/\psi K^{*0}$ control mode is carried out to extract a set of *sWeights* which can then be applied to obtain a signal only data sample from the data. As the $B^0 \to J/\psi K^{*0}$ gives access to a large signal, a clean sample of these events can be extracted with only the preselection, a loose cut to reduce combinatorial background (detailed in Sec. 5.1.4) and some loose particle identification criteria (detailed in Sec. 5.1.5) applied.

At this stage, some known differences between data and simulation can be corrected for. In general the detector occupancy, the vertex χ^2 of the B meson and the p_T of the B meson are known to be mis-modelled in the LHCb simulation. The correct description of these variables is important as the efficiency to select candidates has a large dependence on the kinematics of the final state particles, which can be characterised by the B p_T, and detector occupancy. The distributions of these variables in simulation can be weighted to match the *sWeighted* distributions extracted from data. This is done in an iterative procedure such that first the simulation is weighted to match the occupancy, then to match the B p_T and finally the B vertex χ^2, with the previous steps weights being applied when carrying out the next step. The results of this weighting procedure for both Run 1 and Run 2 data (where slightly different occupancy variables are used due to a bug in the track multiplicity variable in the Run 2 dataset and so SPD multiplicity is used instead) are shown in Fig. 5.2. The same weights derived from the $B^0 \to J/\psi K^{*0}$ mode are also used to correct the simulated samples for the other modes previously mentioned.

It should be noted that other then the different occupancy variable being used in the Run 2 dataset, there is an additional difference in the weighting procedure used. Due to the previously discussed difference in GEC applied on the two L0 trigger lines used in the analysis for Run 2, there is a discontinuity in the SPD multiplicity variable due to the tighter `L0Muon` requirement. This discontinuity is not a detector effect as such and so the difference between simulation and data can not just be simple weighted for. Instead the kinematic weights applied are derived from only events passing through the `L0DiMuon` trigger where the loosest cut is applied. As such, the discontinuity does not effect the derived weights as it is not present in this sample. These weights are then applied to the whole sample. From Fig. 5.2 it is clear that this is sufficient to obtain a good agreement between data

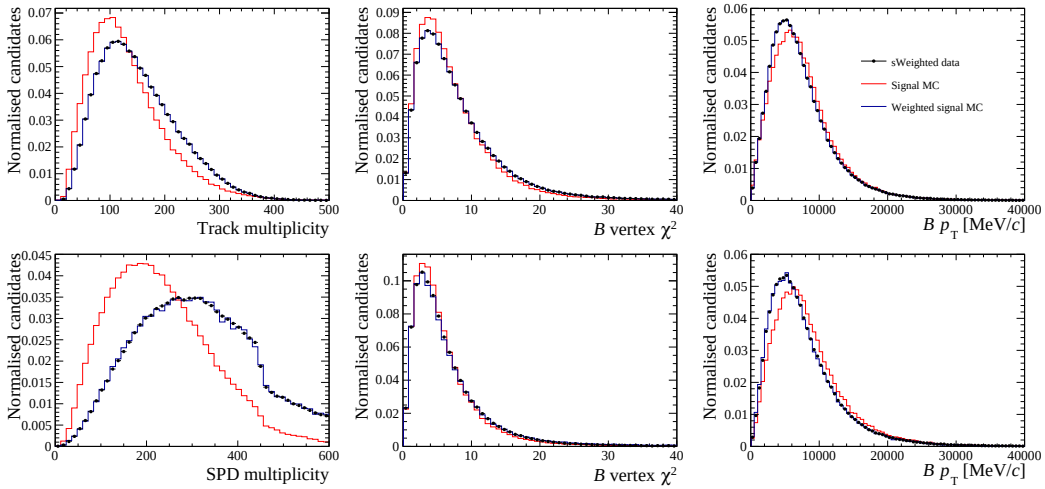

Figure 5.2: Results of the MC weighting procedure for Run 1 (top) and Run 2 (bottom). Histogram areas are all normalised to one. From left the right, variables are shown in the order the iterative weighting procedure is applied.

and simulation.

In addition to the kinematic weighting applied to all decay modes, it is necessary to apply a further correction to the samples of B_s^0 decays. The production ratio of the B^0 and B_s^0 mesons has been measured at the LHCb experiment [126] as a function of B meson transverse momenta and has found to be

$$f_s/f_d = (0.256 \pm 0.020) + (-2.0 \pm 0.6) \times 10^{-3} \times (p_{\mathrm{T}} - 10.4)/\,\mathrm{GeV}/c. \qquad (5.1)$$

This can be compared with the relative fraction of B^0 and B_s^0 in bins of transverse momenta from the simulated samples used in this analysis and weights applied to correct for the difference between the two. This correction is used when calculating the efficiencies of the B_s^0 modes as described in Chapter 6.

5.1.4 Initial multivariate selection

Combinatorial background events, where random tracks in the detector are combined to give a broad background, are primarily handled through the use of an artificial neural network. This allows information from a variety of different variables to be combined into a single variable which can be used to discriminate between signal and background events. By training the neural network using a sample of signal events and a sample of background events, the final neural network can be applied to a mixed sample and used to separate the two.

For this analysis, the NeuroBayes [127] neural network package is used. A

signal sample is provided via a MC sample of $B^0 \to K^{*0}\mu^+\mu^-$ events, chosen over a $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ sample due to the availability of a significantly larger sample of MC events with minimal differences in the relevant variables between the modes. The background sample is taken directly from data, taking events which have passed the trigger and stripping selections in the upper mass sideband (masses greater that $5670\,\mathrm{MeV}/c^2$ and outside of the planned invariant mass fit range).

The general considerations of what input variables to use in the neural network training are to select those where this a large difference between distributions of signal and background events. There is also the requirement that any variable used must be well modelled by the MC samples used in the training. This is the ensure that the neural network distribution in MC matches the true distribution of events in data. The conditions at the different centre of mass energies between Runs 1 and 2 require different neural networks to be trained for each dataset. In addition, different variables of interest are available between the datasets and so the input variables for each network differ slightly. The input variables, as well as their rankings in the neural network and thus relative importance as discriminators in the network, are shown in Tab. 5.3. A comparison of variable distributions from simulation and combinatorial background are shown in Fig. 5.6 and Fig. 5.7 for Run 1 and Run 2 respectively. The data-simulation agreement for these variables is shown in Fig. 5.8 and Fig. 5.9 for Run 1 and Run 2 respectively, after the application of the kinematic weighting described in Sec. 5.1.3.

The variables used have all previously been described in Sec. 5.1.2 with the exception of the isolation variables for each final state track. The isolation variables count the number of tracks not part of the reconstructed candidate in a forward cone from the track. For signal like decays, it is expected that the final state tracks should be well isolated and thus the isolation variable value low. Note that the calculation of the track isolation requires information about other tracks in the event that are not part of a candidate. In Run 2, the datasets containing these candidates was reduced to not contain information about the rest of the event in the interest of preserving disk space and so these isolation variables could not be calculated. A comparison of the performance of the two networks can be carried out by investigating the background rejection rates of the networks versus the signal efficiency. The higher the background rejection rate of the neural network for a given signal efficiency, the better the performance of the variable. This comparison is shown in Fig. **??**, where the area to the left of the curves is a metric of the performance of the networks. From this, it is clear that the Run 1 network performs better than the Run 2 network, this is expected due to the availability of the isolation variables in Run 1.

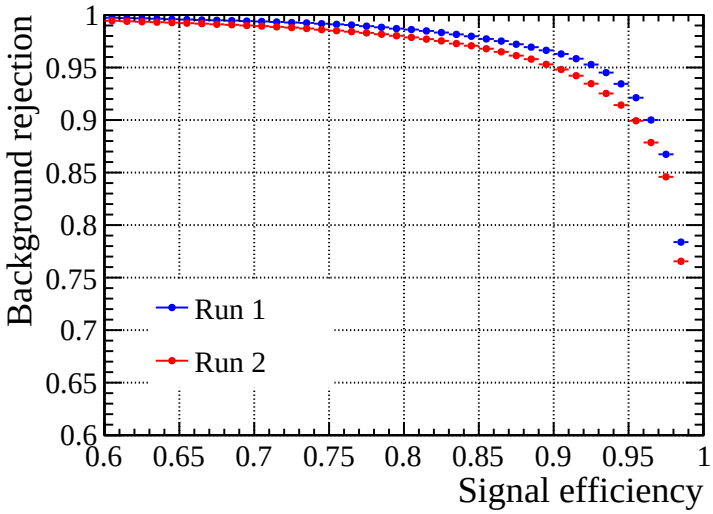

Figure 5.3: Background rejection rate versus signal efficiency for the Run 1 (blue) and Run 2 (red) neural networks.

Table 5.3: Input variables and their relative importances in the neural network training for Run 1 and Run 2 datasets. Where no number is given, the variable is not included in the training for that dataset.

Particle	Variable	Ranking	
		Run 1	Run 2
B	vertex χ^2	1 (best)	1 (best)
	cos(DIRA angle)	10	9
	χ^2_{IP}	5	5
	Lifetime	12	6
	p_{T}	-	4
$\mu^+\mu^-$	vertex χ^2	13 (worst)	10 (worst)
	flight distance χ^2	3	2
	χ^2_{IP}	11	8
K	Isolation	6	-
π	Isolation	9	-
μ^+	Isolation	7	-
μ^-	Isolation	2	-
Daughter particles	Minimum p_{T}	4	3
	Maximum p_{T}	8	7

After training the neural network and applying it to the data sample, the output variable provides a good discriminant between signal and background. Fig-

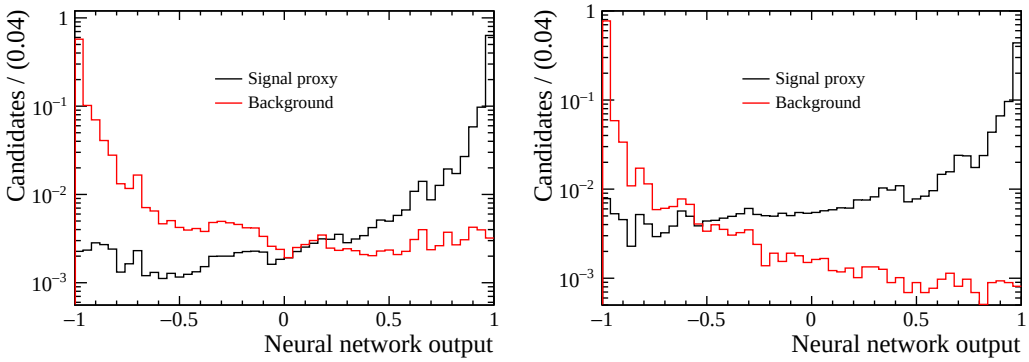

Figure 5.4: Neural network responses for both simulated signal decays and combinatorial background events for Run 1 (left) and Run 2 (right) data.

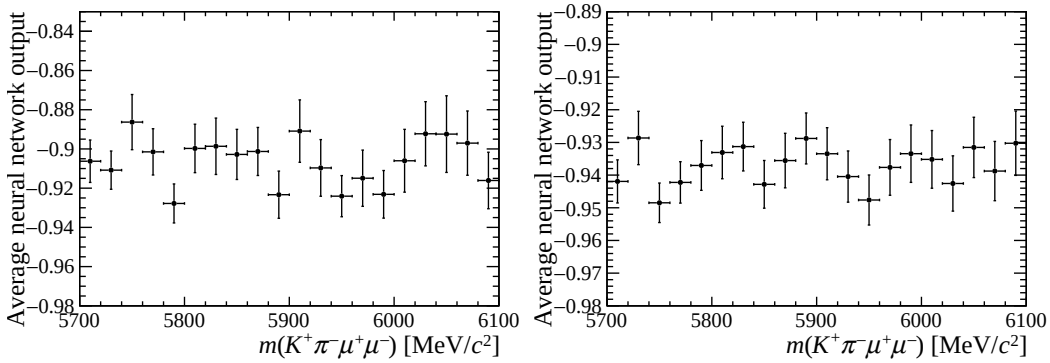

Figure 5.5: Average neural network responses in bins of candidate invariant mass for combinatorial background events extracted from data for Run 1 (left) and Run 2 (right).

ure 5.4 shows a comparison of the output variable taken from the simulated signal sample and the background sample for both Runs 1 and 2. The neural network is investigated for overtraining by checking that there is no correlation between candidate mass and neural network output. The average neural network output in bins of candidate mass for combinatiorial background for both the Run 1 and Run 2 network is shown in Fig. 5.5 and no correlation is seen. It is also checked that the output on simulation agrees with the output from data, making use of the same $sWeights$ and $B^0 \to J/\psi K^{*0}$ sample used in the kinematic weighting. This is necessary to ensure that efficiencies of the neural network taken from simulation are in agreement with the true efficiency. At this point, a loose cut on the neural network is applied that is 95% efficient on the signal sample and that rejects approximately 90% of background events. The further optimisation of the neural network is discussed in Sec. 5.3.

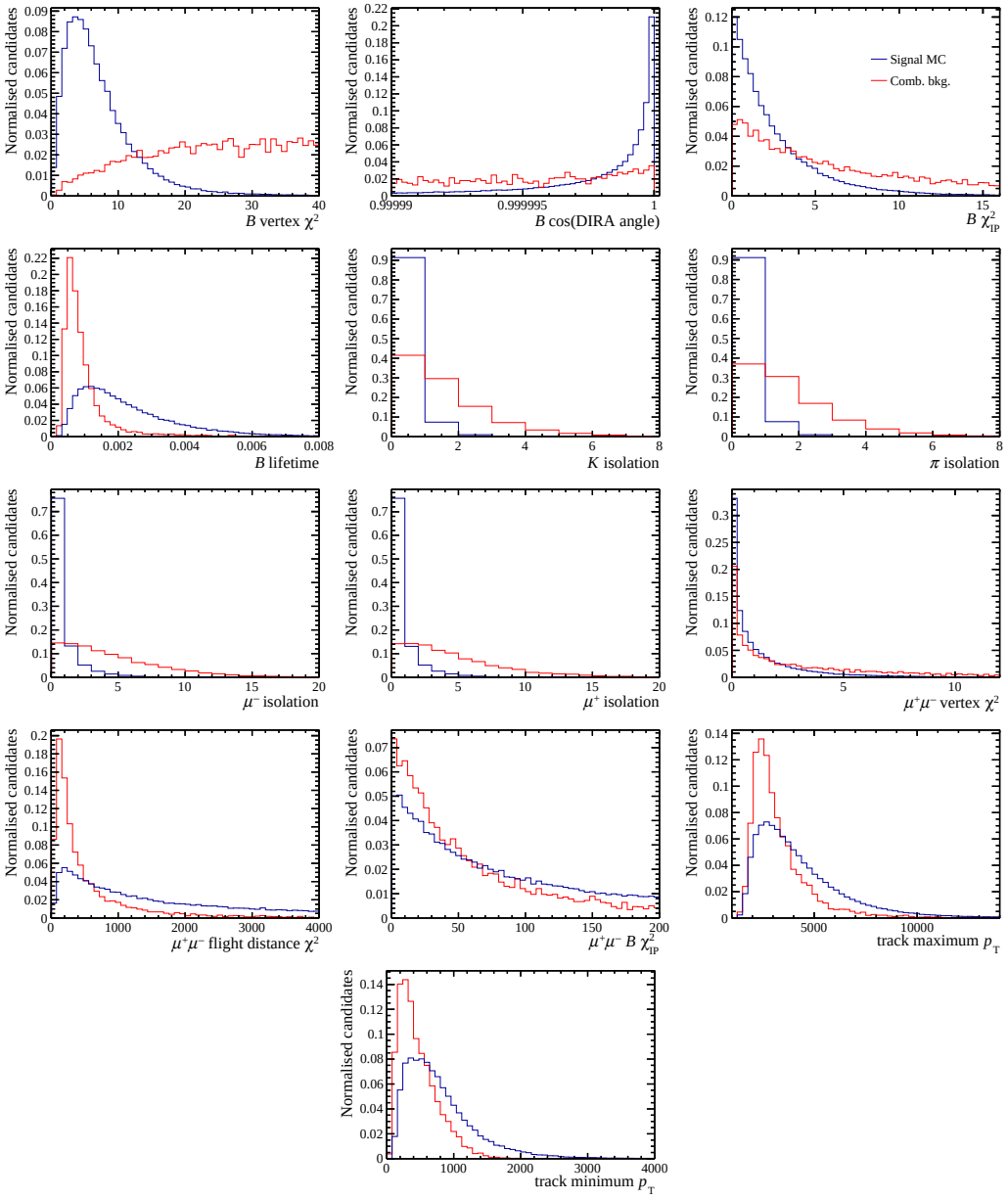

Figure 5.6: Input variables for Run 1 neural network training. Blue is the $B^0 \to K^{*0}\mu^+\mu^-$ signal distribution from simulation, red is the combinatorial background distribution from the upper mass sideband of data. Histogram areas are normalised to one.

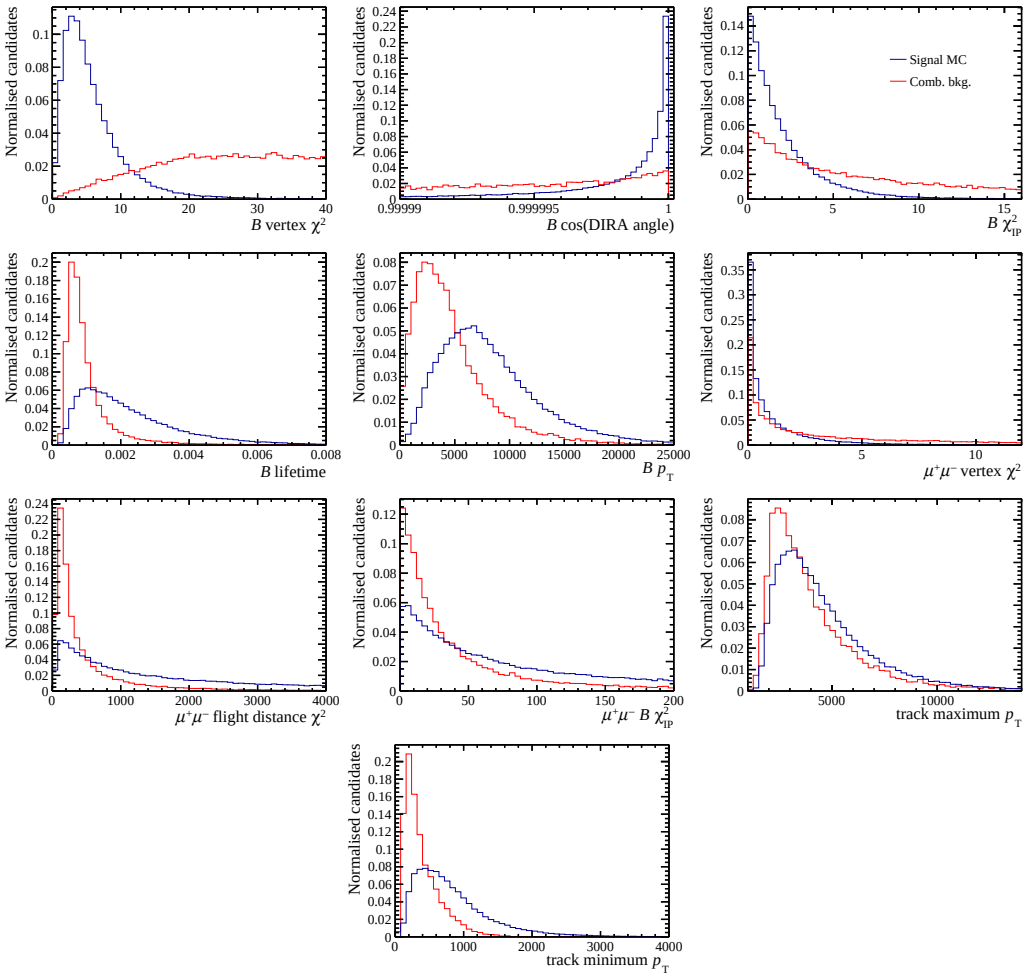

Figure 5.7: Input variables for Run 2 neural network training. Blue is the $B^0 \to K^{*0}\mu^+\mu^-$ signal distribution from simulation, red is the combinatorial background distribution from the upper mass sideband of data. Histogram areas are normalised to one.

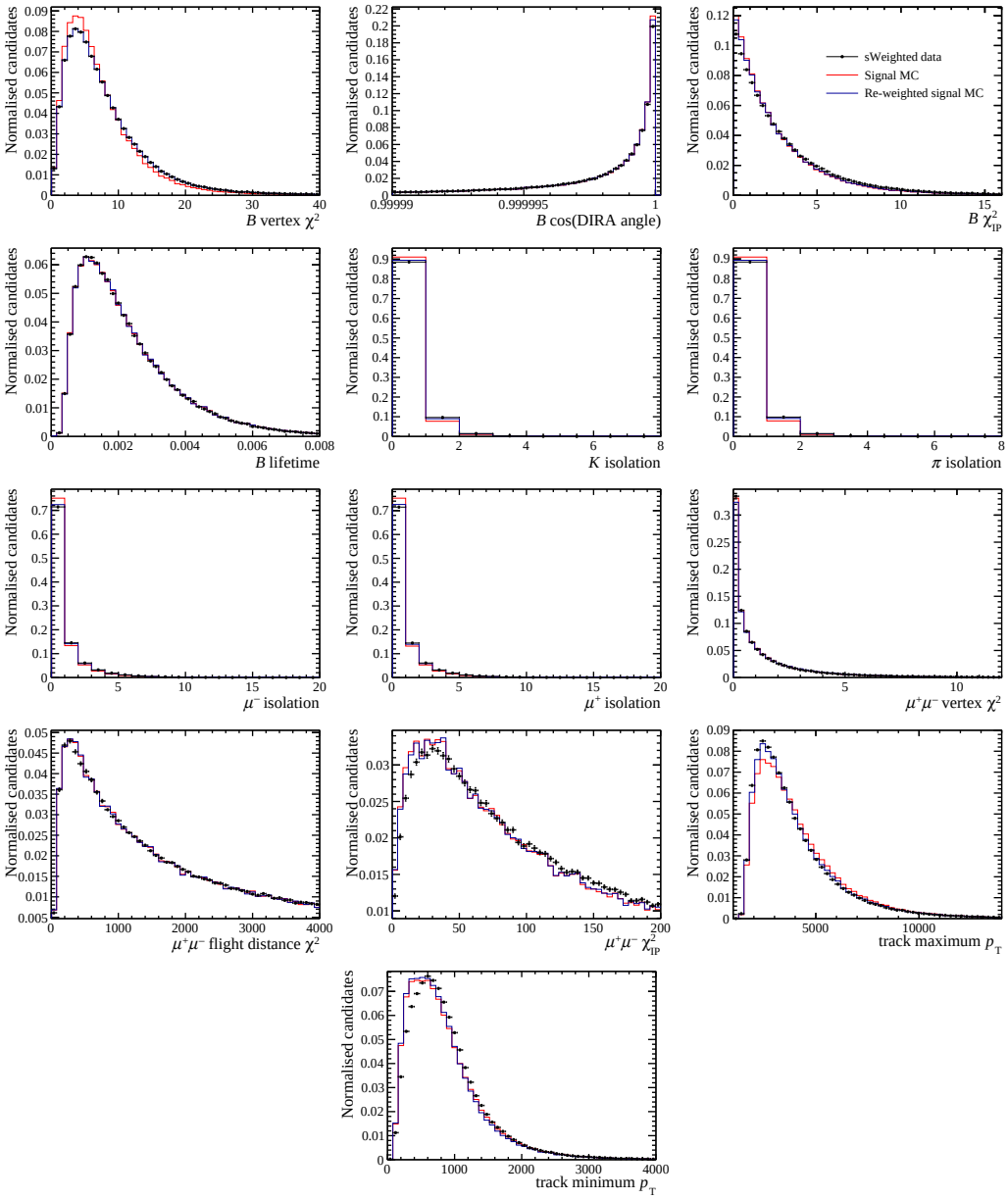

Figure 5.8: Data-simulation agreement for the input variables for the Run 1 neural network training. Red is the uncorrected $B^0 \to J/\psi K^{*0}$ distribution from simulation, blue is the corrected $B^0 \to J/\psi K^{*0}$ signal distribution from simulation, and black points are from a *sWeighted* $B^0 \to J/\psi K^{*0}$ extracted from data. Histogram areas are normalised to one.

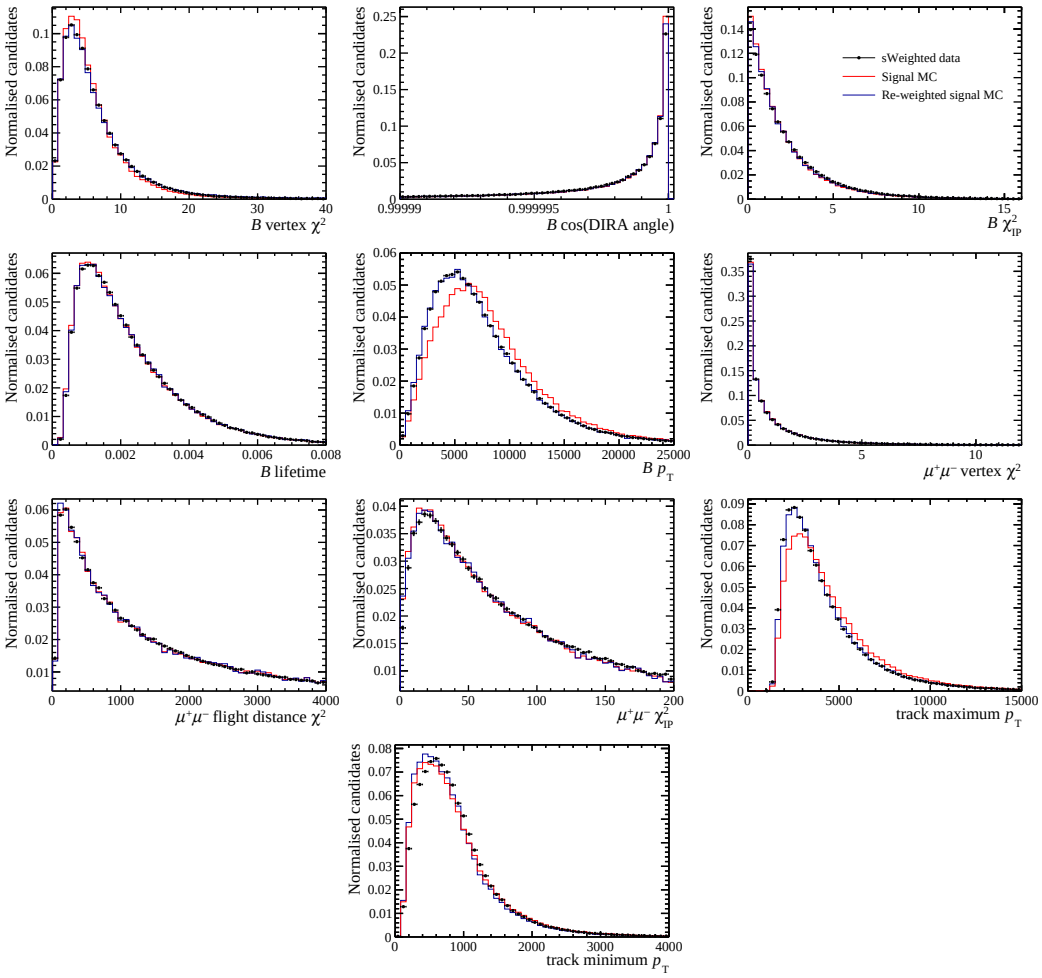

Figure 5.9: Data-simulation agreement for the input variables for the Run 2 neural network training. Red is the uncorrected $B^0 \to J/\psi K^{*0}$ distribution from simulation, blue is the corrected $B^0 \to J/\psi K^{*0}$ signal distribution from simulation, and black points are from a *sWeighted* $B^0 \to J/\psi K^{*0}$ extracted from data. Histogram areas are normalised to one.

5.1.5 Particle identification variables

As discussed in Sec. 3.2.9, LHCb makes use of a combination of information from different sub-detectors to provide the best discriminating variables between different particle types. Here the variables that make use of multivariate analysis techniques, the `ProbNN` variables, are used. These variables take values between 0 and 1, with an increasing value indicating a greater probability that a particle is a given species depending on the `ProbNN` variable in question.

For the final state hadrons considered, a combination of `ProbNN` variables is used, taking into account the likelihood that a given particle is the species it has been reconstructed as and the likelihood that it is not another specific species. As kaon-pion misidentification is the most likely at LHCb, initial particle identification requirements of $\texttt{ProbNN}_\pi \times (1 - \texttt{ProbNN}_K) > 0.05$ for the pion and $\texttt{ProbNN}_K \times (1 - \texttt{ProbNN}_\pi) > 0.05$ for the kaon are applied. This initial requirement removes a large amount of background events providing a far more manageable dataset in which to investigate any additional backgrounds. Further optimisation of these variables, taking into account known backgrounds in the datasets, is discussed in Sec. 5.4.

5.1.6 Signal resolution

Due to the large number of expected B^0 events compared to B_s^0 events in the fit, any improvement of the signal invariant mass resolution is important to try and maximise the sensitivity of the fit to the $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ decay. Studies are carried out to determine the signal mass resolution dependence on two sets of variables: the uncertainty on the measured mass of reconstructed candidate and the opening angles between pairs of reconstructed tracks. These studies are described in the following section.

Measured mass uncertainty

The measured mass uncertainty is estimated as part of the kinematic fit when reconstructing a candidate, with the uncertainty on track momentum and the opening angle between tracks defining the mass uncertainty. Putting an upper limit on the allowed measured mass uncertainty of reconstructed candidates should improve the invariant mass resolution. Figure 5.10 shows both the resolution and the relative efficiency for cuts on this variable. While requiring candidates have a low measured mass uncertainty clearly does improve the resolution, there is also a correspondingly large drop in signal efficiency. A loose requirement that candidates have a measured

mass uncertainty less than $22\,\mathrm{MeV}/c^2$ is applied which is 95% efficient on the signal. This has a minimal effect on the core resolution of the signal but does remove candidates from the tails of the signal at high invariant mass, which is important for control of the tails of the $K^+\pi^-\mu^+\mu^-$ invariant mass distribution that will be further discussed in Sec. 7.1.1.

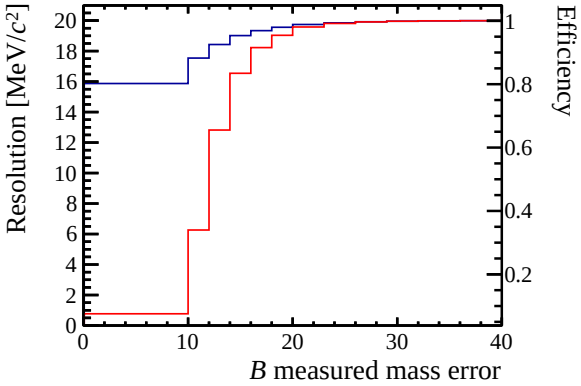

Figure 5.10: Resolution and efficiency as a function of an upper limit on measured mass uncertainty taken from simulated $B^0 \to K^{*0}\mu^+\mu^-$ candidates. The blue line shows the resolution (taken as the RMS of $K^+\pi^-\mu^+\mu^-$ invariant mass from simulation) and the red line shows the relative efficiency of applying the cut.

Angles between tracks

In a similar manner, requiring a larger angular separation between pairs of final state tracks will improve the invariant mass resolution. The dependence on track opening angle is investigated by requiring that the angular separation between either the final state kaon and pion or between the final state muons is greater than a specified value and taking the resolution as the RMS of $K^+\pi^-\mu^+\mu^-$ invariant mass from simulated $B^0 \to K^{*0}\mu^+\mu^-$ candidates for a range of angular separations. Figure 5.11 shows the results of this study. As clearly demonstrated in the figure, while there is a gain in resolution when applying requirements on these variables, the efficiency drop for a given resolution gain is even larger than that observed when applying a similar procedure on the measured mass uncertainty. The loss in signal efficiency for a given gain in resolution is considered to be too high and so only a loose requirement on these angular separations is applied, primarily to remove a background where a track has been cloned in the reconstruction. Figure 5.12 shows the angular separation between all possible two track combinations of the final state particles, clearly indicating the peak at low angular separations that is removed with

the requirement that all angular separations must be greater than 0.0005.

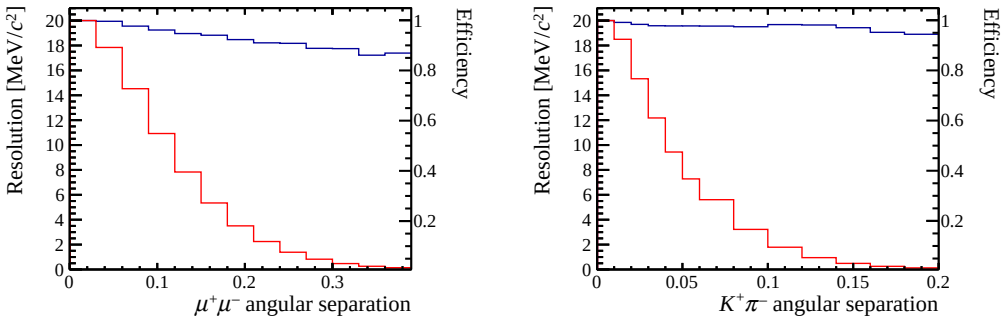

Figure 5.11: Resolution and efficiency as a function of a lower limit on angular separation variables calculated from simulated $B^0 \to K^{*0}\mu^+\mu^-$ candidates. In each plot, the blue line shows the resolution (taken as the RMS from simulation) and the red line shows the relative efficiency of the angular separation requirement. The angular separations shown in each plot are between $K^+\pi^-$ (left) and $\mu^+\mu^-$ (right).

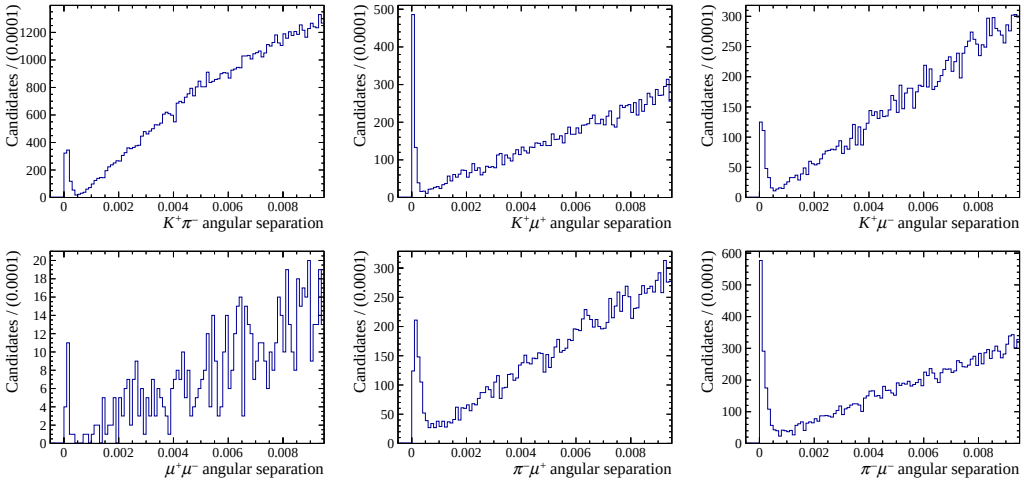

Figure 5.12: Distributions of angular separations between all two track combinations of the final state particles for candidates extracted from data.

5.2 Backgrounds

In any search for a decay mode, it is of great importance to ensure that any specific backgrounds are well considered and understood. The multivariate selection described in Sec. 5.1.4 removes background events where tracks have been mistakenly combined, however this will not reject backgrounds with a similar topology to the searched for signal. For this highly suppressed mode it is essential that all backgrounds that are present in the signal region are well understood and that a maximum sensitivity to the signal is preserved. In this section, a detailed overview of the potential physics backgrounds, whether they are expected to have a significant impact in the invariant mass window near to the expected signal and what action is taken to minimise any impact is given.

Two general approaches to specific backgrounds are taken in this analysis. The first is to apply vetoes that efficiently remove background events while minimising any impact on the signal efficiency. Generally these vetoes are applied in a specific mass region associated with the background and using a discriminating variable between the background and the signal. In some cases, a veto would negatively impact the signal efficiency to a large degree. In these instances a second approach of including a component for the background in the invariant mass fit model is taken. For each of the backgrounds presented below either the veto that is used or the method in which a background component is determined are detailed.

5.2.1 Fully hadronic backgrounds

Backgrounds from fully hadronic four body final state decays, where two of the hadrons are misidentified as muons, are heavily suppressed due to the `IsMuon` requirement in the `B2XMuMu` stripping line. To investigate the presence of any residual backgrounds, the $\texttt{ProbNN}_{\mu}$ variable can be taken directly from data to see if there is any remaining background at low values, the results of which are shown in Fig. 5.13. In this figure, it is clear that there is still a significant number of events where the reconstructed muon has a very low probability to be a muon. As such a requirement that $\texttt{ProbNN}_{\mu} > 0.05$ for both reconstructed muons is applied, this cut is indicated in the plot.

To determine the potential impact of any hadronic backgrounds that remain after the applied particle identification requirements, the effective branching fractions for the most likely background modes can be calculated. These branching fractions are calculated relative to the $B^0 \to K^{*0}\mu^+\mu^-$ branching fraction and therefore include a scaling due to f_s/f_d for decays of B^0_s mesons as well as a scaling

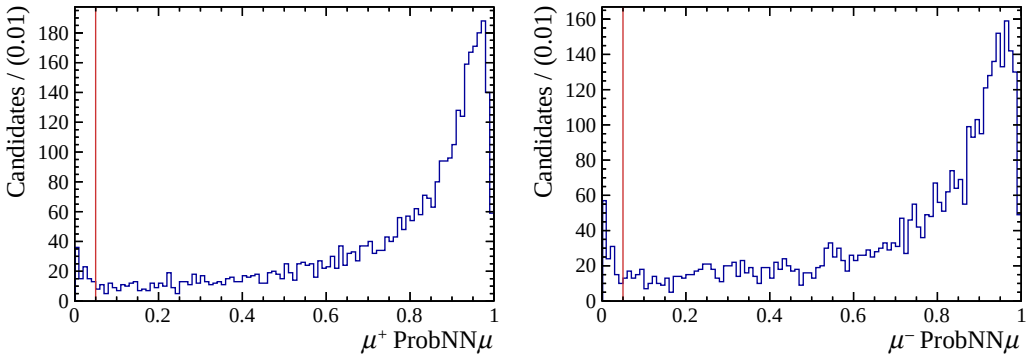

Figure 5.13: ProbNNμ distributions of selected $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ events for both final state muons from data. The location of the selection criteria applied to these variables is indicated by the dashed red line.

from particle misidentification rates. To calculate the misidentification rate of kaons or pion to muons in the LHCb detector the PIDCALIB package [128] is used. This makes use of calibration samples from pure decay modes where the species of the final state particles are well known. The efficiencies of a specific PID cut can then be evaluated on these pure samples in different kinematic regions and this information used to calculate the efficiency for a given particle track. The misidentification rate can then be calculated on a mode by mode basis, taking into account the `IsMuon` and $\texttt{ProbNN}_\mu$ criteria. Table 5.4 summarises these decay modes, their measured branching fractions and effective branching fractions.

Given that the predicted effective branching fraction for the $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ mode is approximately 7×10^{-9} including the B_s^0 production scaling, the effective branching fraction of any potential background must be $\mathcal{O}(10^{-11})$ or lower to be considered negligible. The decays $B^0 \to K^{*0}(892)\pi^+\pi^-$ and $B^0 \to K^{*0}(892)K^+K^-$ have effective branching fractions at the expected level of the signal decay but sit outside the B_s^0 mass window, under the signal from the much larger $B^0 \to K^{*0}\mu^+\mu^-$ decay. Therefore it is the comparison to the $B^0 \to K^{*0}\mu^+\mu^-$ that is important for these decays and as they have effective branching fractions that are below 1% of this mode they can be safely considered negligible. For background decays of B_s^0 mesons, the limit on the branching fraction of $B_s^0 \to K^{*0}(892)\rho^0$ indicates that this background could be at the level of the signal decay. However, comparing this decay to similar processes for the B^0 ($B^0 \to \rho\rho$ and $B^0 \to \rho\omega$) results in an branching fraction estimate that is three orders of magnitude smaller than the current limit. As such, this background can also be safely neglected. The largest remaining fully hadronic background in the signal region is that from $B_s^0 \to K^{*0}(892)K^{*0}(892)$ with an effective branching fraction that is approximately 2% of the expected signals. Any

potential effect of this background will be considered as a systematic uncertainty and discussed further in Chapter 8.

Table 5.4: Possible fully hadronic decay backgrounds and their branching fractions taken from Ref [2] except $B_s^0 \to \overline{K}^{*0}(892)K^{*0}(892)$ which is updated from Ref [129]. Effective branching fractions are calculated using PIDCalib and the LHCb value for f_s/f_d for B_s^0 decays.

Decay mode	Branching fraction	Effective branching fraction
$B^0 \to K^{*0}(892)\pi^+\pi^-$	$(5.5 \pm 0.5) \times 10^{-5}$	$(2.48 \pm 0.23) \times 10^{-9}$
$B^0 \to K^{*0}(892)K^+K^-$	$(2.75 \pm 0.26) \times 10^{-5}$	$(2.14 \pm 0.20) \times 10^{-9}$
$B_s^0 \to K^{*0}(892)\rho^0$	$< 7.67 \times 10^{-4}$	$< 8.34 \times 10^{-9}$
$B_s^0 \to \overline{K}^{*0}(892)K^{*0}(892)$	$(1.08 \pm 0.41) \times 10^{-6}$	$(1.5 \pm 0.6) \times 10^{-10}$
$B_s^0 \to \phi K^{*0}(892)$	$(1.13 \pm 0.30) \times 10^{-6}$	$(2.2 \pm 0.6) \times 10^{-11}$

5.2.2 B to charm decays

Decays involving a $b \to c$ transition are relatively common, especially with respect to the rare decay mode that is of interest in this analysis. As such a variety of decays of this type can constitute backgrounds, including semileptonic decays of the type $b \to c(\to s\mu\nu X)\mu\nu Y$. These can then be reconstructed, missing the neutrinos, and will form a part of the background at low reconstructed B invariant masses. The relatively tight requirement on the reconstructed $K^{*0}(892)$ invariant mass means that backgrounds with a real K^{*0} are heavily favoured.

A search for background events of this type can be carried out by searching for peaking structure at the $D^0/D^\pm$ mass when reconstructing the $K^+\mu^-$ or $K^+\pi^-\mu^-$ invariant masses. The results of this reconstruction are shown in Fig. 5.14 and clearly show indications of background of this type. The simplest method of removing this background is to veto events which peak at the known $D^0/D^\pm$ mass under this reconstruction. Any events with a reconstructed invariant mass in the range $1840 < m(K^+\pi^-\mu^-) < 1880\,\text{MeV}/c^2$ or $1840 < m(K^+\mu^-) < 1880\,\text{MeV}/c^2$ are removed. As these events will sit at lower reconstructed B^0 invariant mass than the expected signal this veto is very efficient on the signal and control mode.

5.2.3 Peaking backgrounds

A number of specific decay modes can constitute backgrounds in the region of the invariant mass fit with a peaking structure which is significantly different from the

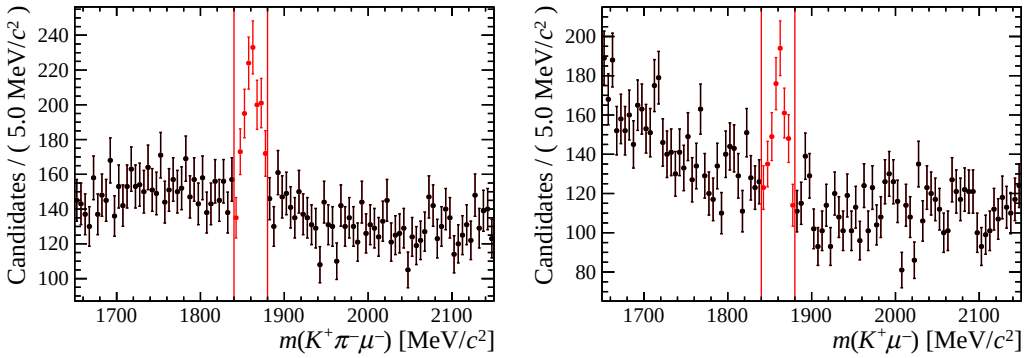

Figure 5.14: Events reconstructed as $K^+\pi^-\mu^-$ (left) and $K^+\mu^-$ (right) demonstrating the presence of charm backgrounds and the effect of applying vetoes to remove these events. Red events are before application of veto, black after application of veto and red lines indicate the invariant mass boundaries of the vetoes.

broad combinatorial background events. Decays of this type generally involve another four body decay that has been fully reconstructed but where there has been some final state particle misidentification on the hadrons. Due to particle identification requirements already applied these tend to be suppressed, especially when there have been multiple misidentifications.

The general strategy to evaluate the impact of these backgrounds is to perform a search for these events in the datasets after the pre-selection is applied. These searches are carried out by changing the mass hypotheses of the final state particles, reconstructing these events and then searching for peaking structure at the invariant mass of the expected background decay. Particle identification requirements can be applied to specifically select out these backgrounds to make the presence of the backgrounds clear. The peaking backgrounds considered in this analysis are detailed in the following sections.

Misidentified kaon $\leftrightarrow$ pion

Events from real $B^0 \to K^{*0}\mu^+\mu^-$ events can act as a background in the analysis if the final state kaon is misidentified as a pion and vice-versa. This results in a broad background centred at the correct B^0 mass with a vastly different shape from signal events that have been correctly reconstructed and which cannot be effectively included in a signal model. To search for this background, the $B^0 \to J/\psi K^{*0}$ control mode is used as the presence of the K^{*0} in the final state leads to a similar misidentification effect. The kaon is reconstructed under the pion mass hypothesis and vice versa and then particle identification cuts of $\texttt{ProbNN}_K \times (1 - \texttt{ProbNN}_\pi) > 0.05$ for the pion and $\texttt{ProbNN}_\pi \times (1 - \texttt{ProbNN}_K) > 0.05$ for the kaon are applied.

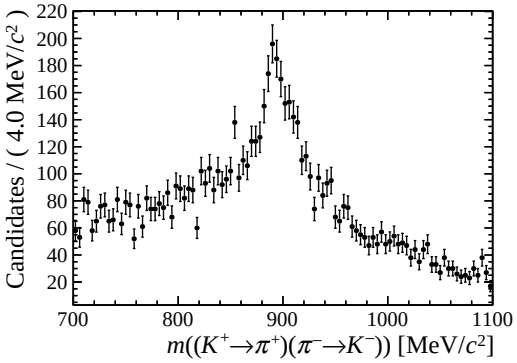

Figure 5.15: Reconstructed $B^0 \to J/\psi K^{*0}$ decays under the $K \leftrightarrow \pi$ mass hypothesis and with PID requirements to select the mis-ID component.

The results of carrying out the search for this background are shown in Fig. 5.15 which clearly demonstrates a peaking structure in $m((K^+ \to \pi^+)(\pi^- \to K^-)$ on a broad background. Note that here and in the following sections, the notation $p_1 \to p_2$ represents when the invariant mass has been recalculated under the mass hypothesis that a final state particle that was originally reconstructed as species p_1 is actually species p_2. This background is naturally suppressed by the application of particle identification criteria on the pion and kaon however a small contribution from these decays still remains. A component in the invariant mass fit which takes into account the suppression from particle identification requirements is included for these decays, the details of which are more thoroughly discussed in Sec. 7.1.

Misidentified hadron $\leftrightarrow$ muon

Events from true $B^0 \to J/\psi K^{*0}$ and $B^0 \to \psi(2S)K^{*0}$ decays can constitute a background in the $B^0 \to K^{*0}\mu^+\mu^-$ mode when a muon is identified as one of the final state hadrons and vice-versa. These events are shifted in dimuon invariant mass and thus are no longer removed by the vetoes in q^2. A search for these events is carried out by reconstructing the $K^+(\pi^-)$ as the $\mu^+(\mu^-)$ and vice-versa. Any events where the new reconstructed dimuon invariant mass peaks at the known masses of the J/ψ and the $\psi(2S)$ can by identified as background events.

The mass reconstructions under the double swapped mass hypothesis for both the kaon $\leftrightarrow$ muon and pion $\leftrightarrow$ muon swaps are shown in Fig. 5.16 in both the J/ψ and the $\psi(2S)$ mass regions. These events can be cleanly vetoed from the dataset by making use of the `IsMuon` information for the final state hadrons. Events are vetoed if they are consistent with the J/ψ ($3036 < m(K/\pi \to \mu)\mu < 3156\,\mathrm{MeV}/c^2$)

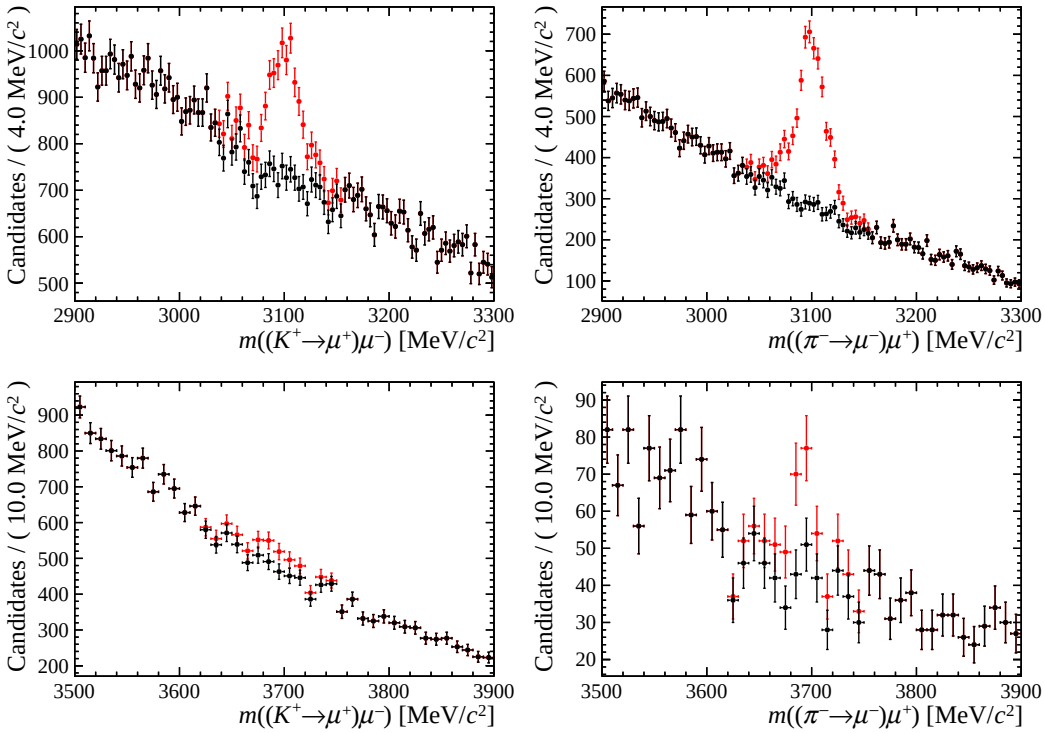

Figure 5.16: Events reconstructed under the double-swapped mass hypotheses $K \leftrightarrow \mu$ (left) and $\pi \leftrightarrow \mu$ (right) in the J/ψ (top) and $\psi(2S)$ (bottom) mass regions. Red events are before application of veto, black after application of veto.

or the $\psi(2S)$ ($3626 < m(K/\pi \to \mu)\mu < 3746\,\text{MeV}/c^2$) mass hypotheses and if the kaon or pion does not satisfy the `IsMuon` criteria. The effects of these vetoes are also shown in Fig. 5.16.

$\Lambda_b^0 \to pK^-\mu^+\mu^-$

The baryonic decay mode $\Lambda_b^0 \to pK^-\mu^+\mu^-$ can act as a background in this analysis when the final state proton is mis-identified as a pion. This is of particular concern due to the background peaking in the same mass region under the standard mass reconstruction as the rare signal being searched for. Understanding and controlling the contribution from this background is a key part of maximising the sensitivity to the signal mode.

To get the best understanding of this background, a sample of the background can be extracted from the dataset. A search can be carried out by reconstructing the mass under the mass hypothesis that the final state pion is a proton and with the particle identification cut $\texttt{ProbNN}_p \times (1-\texttt{ProbNN}_K) \times (1-\texttt{ProbNN}_\pi) > 0.05$ for the pion to select out this background. The results of this search on the reconstructed

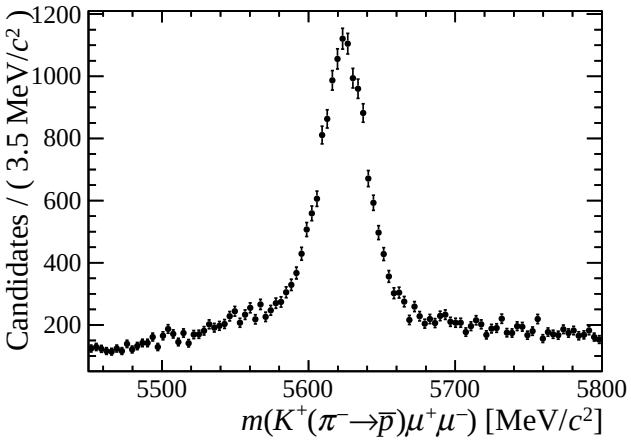

Figure 5.17: Events reconstructed under the $\pi \to p$ mass hypothesis in the Λ_b^0 mass region. Loose PID requirements have been applied to select out this misidentified component.

data after the application of the initial selection requirements is shown in Fig. 5.17, where the control mode is used for the search. It is clear that there is a significant contribution from $\Lambda_b^0 \to J/\psi p K^-$ events in the dataset and thus it can be expected that $\Lambda_b^0 \to p K^- \mu^+ \mu^-$ events will contribute in the rare mode. One way of dealing with this background would be the application of a veto to remove these events from the dataset. However, due to the fact that this background sits in the same invariant mass region as the searched for signal, the application of such a veto would have to be inefficient on the signal to cleanly remove the background. To preserve as much of the searched for signal as possible, a veto is not applied and instead a component for this background is included in the invariant mass fit model. The method by which a line shape for this background is extracted and how the yield of this background is constrained is discussed in Sec. 7.1.4.

$B_s^0 \to \phi\mu^+\mu^-$

Another peaking background which can contribute to this analysis and only requires a single hadron misidentification, a pion misidentified as a kaon, is that from $B_s^0 \to \phi\mu^+\mu^-$. A search is performed by reconstructing the B meson with the final state pion substituted with a kaon. The narrow ϕ meson state provides a very clean experimental signature and as such does not require any additional particle identification to select out. The reconstructed ϕ mass in the control mode dataset is shown in Fig. 5.18, with a clear peak at the $\phi(1020)$ mass.

Unlike the background from $\Lambda_b^0 \to p K^- \mu^+ \mu^-$, this background does not peak in the $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ signal region. As such, a veto to remove this decay

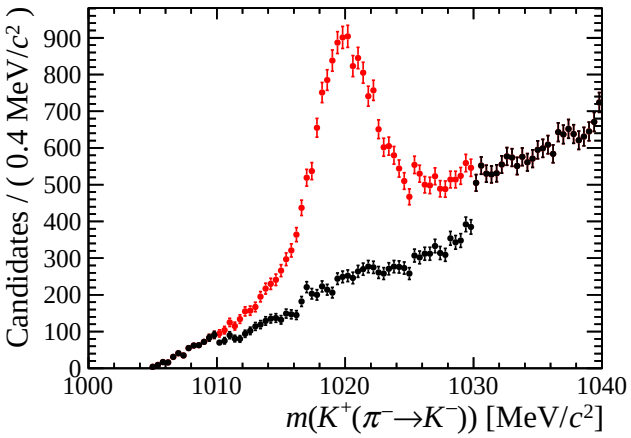

Figure 5.18: Events reconstructed under the $\pi \to K$ mass hypothesis in the $\phi(1020)$ mass region of the control mode. Red events are before application of veto, black after application of veto.

mode does not significantly affect the signal efficiency and can thus be applied without removing signal events. A veto making use of the narrow mass region that the background occupies and particle identification information for the particle normally reconstructed as a pion is applied. Events that fall within the mass region $1010 < m(K^+(\pi^- \to K^-)) < 1030\,\text{MeV}/c^2$ and that have a pion with $\texttt{ProbNN}_K > 0.05$ are removed from the dataset. This has an efficiency of over 99.9% on B_s^0 signal candidates while cleanly removing the $B_s^0 \to \phi\mu^+\mu^-$ background. The effect of this veto is also shown in Fig. 5.18.

$B^0_{(s)} \to \pi^+\pi^-\mu^+\mu^-$

Just as there can be misidentification of a kaon as a pion leading to the presence of the above background, a pion can be misidentified as a kaon. Events with a $\pi\pi\mu\mu$ final state come primarily from two decay modes, $B^0 \to \rho\mu^+\mu^-$ and $B_s^0 \to f^0\mu^+\mu^-$. Unlike the $\phi(1020)$, the ρ and f^0 resonances are broad and so directly searching for these resonances under the misidentified mass hypothesis is unfeasible. Instead a search for peaking structure around the B^0 or B_s^0 mass is carried out, the results of which are shown in Fig. 5.19 including the control mode. Peaks are only visible when the initial PID requirements described in Sec. 5.1.5 are not applied, after the application of the initial PID criteria there is no indication of this background in the control mode. As the optimised PID requirements will be tighter than the ones applied here, this background is assumed to be negligible.

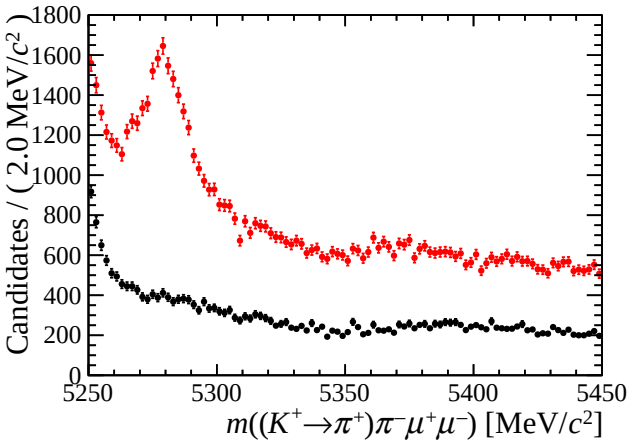

Figure 5.19: Reconstructed B mass under the $K \to \pi$ mass hypothesis. Red events are before application of initial particle identification criteria, black events after.

5.2.4 Mis-reconstructed decays

A number of backgrounds fall into the broad category of mis-reconstructed decays. This refers to decays which constitute backgrounds due to a mistake in the reconstruction that is not related to a mistaken identification of a particle, for example the accidental inclusion of a random particle to a decay mode that causes it to emulate the searched for final state. The three types of mis-reconstruction that are considered here are as follows:

- *Partially reconstructed decays*: This refers to backgrounds with final states that have more particles then that of the searched for signal (*i.e.* four final state particles) that have then been reconstructed missing some of these particles.
- *Randomly replaced particles*: When one of the true final state particles from the signal mode has been randomly replaced with the same species of particle from the background.
- *Over reconstructed decays*: Where a background mode with less final state particles then that of the signal mode has had a random particle from the combinatorial background included.

Partially reconstructed decays

Partially reconstructed decays of b-hadrons with a missing hadron are the most common type of partially reconstructed background that can affect this analysis. A limit on the mass that these background will sit at can be estimated by taking the

difference between the mass of the B^+/B^0 mesons and that of a pion, the lightest possible missing particle for these decays. All of these backgrounds will be present at reconstructed masses of less than approximately $5170\,\text{MeV}/c^2$ and therefore this is taken as the lower mass boundary over which the invariant mass fit will be carried out, excluding these partially reconstructed backgrounds.

Randomly replaced particles

The most common type of particle to be randomly replaced in the $K^+\pi^-\mu^+\mu^-$ final state is the pion, there being a large number of pions in the background of the event. When this replacement occurs, the invariant mass shape of the background is difficult to predict as it is not determined from the kinematics of a single b-hadron decay but is instead reliant on the kinematics of a random unrelated pion. To investigate the shape and impact of this background a sample must be extracted from data.

A background sample is extracted by selecting candidates from the dataset which have the same event number and where the final state kaon and muons are common between candidates. This selects out candidates where at least two possible reconstructions have been carried out, one where the true pion from the B meson has been reconstructed and others where a random pion has been selected. The most likely signal candidate for each pair of candidates is then selected based on two criteria: if one candidate is within $\pm100\,\text{MeV}/c^2$ of the B^0 mass and the other is not the first candidate is classified as the signal; if both candidates are within $\pm100\,\text{MeV}/c^2$ of the B^0 mass then the candidate with the highest neural network response is classified as the signal.

The distribution of multiple candidate events extracted from data is shown in Fig. 5.20, with the datasets separated according to the above criteria. A total of 808 events are selected over the entire q^2 range and so the number of events in the rare mode is very low. Although this method will not select out all events of this type, it can be reasonably expected to not be larger by a significant amount. In addition, the broad shape of the background is similar to that from generic combinatorial background with very little peaking structure. The few background events of this type will be accounted for as part of the combinatorial background in the invariant mass fit.

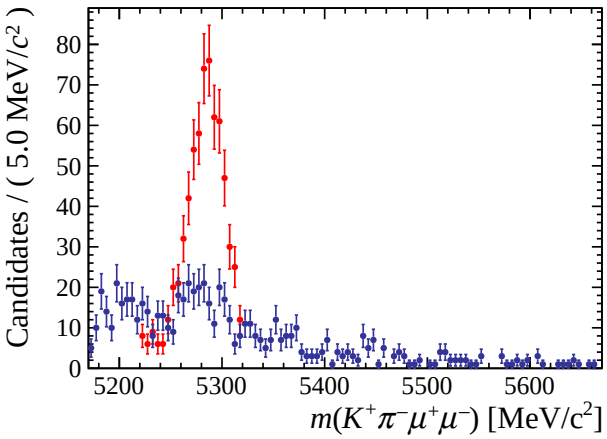

Figure 5.20: Distribution of candidates that share the same event number taken from data. Candidates shown in red are those that are classified as signal whereas candidates shown in blue are classified as background.

Over reconstructed decays

The decay $B^+ \to K^+\mu^+\mu^-$ can act as a background when it has been over reconstructed, *i.e.* one of the large number of random pions in the background has been mistakenly added during the event reconstruction. This decay can be searched for much more easily than the above case with random replacement of a particle as all the particles from the true decay are present for the relevant candidates. Background from $B^+ \to K^+\mu^+\mu^-$ (or $B^+ \to J/\psi K^+$ in the control mode) can be identified by reconstructing $m(K^+\mu^+\mu^-)$ from data and searching for candidates which peak at the B^+ mass. The result of this mass reconstruction over the full q^2 range (including the rare mode and the control mode) is shown in Fig. 5.21 and clearly demonstrates the presence of this background.

The $B^+ \to K^+\mu^+\mu^-$ background is located at higher invariant masses then the signal mode due to the addition of a random pion. As such, this background is outside of the sensitive invariant mass region where the rare signal is expected and so a veto on the background will not negatively effect signal efficiency. However the application of such a veto would apply a "sculpting" effect to the combinatorial background. Figure 5.21 shows the distribution of the background both in the search reconstruction, $m(K^+\mu^+\mu^-)$, and in the the standard signal reconstruction, $m(K^+\pi^-\mu^+\mu^-)$. From this it is clear that by removing the $B^+ \to K^+\mu^+\mu^-$ events with a veto on $m(K^+\mu^+\mu^-)$ proportionally more combinatorial background is removed at high invariant mass then there is at low invariant mass. This removes background candidate in the sideband that are not removed from the signal region

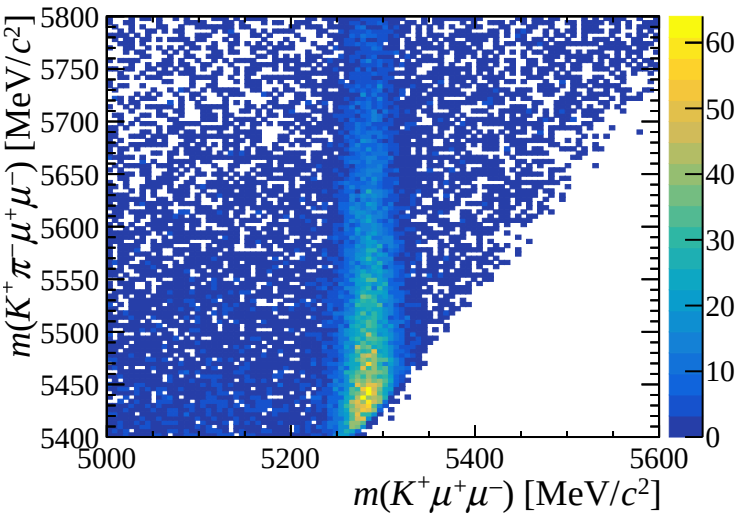

Figure 5.21: Distribution of candidates in the upper mass sideband in both $m(K^+\pi^-\mu^+\mu^-)$ and $m(K^+\mu^+\mu^-)$. The vertical band corresponds to background from $B^+ \to K^+\mu^+\mu^-$ decays (where the J/ψ and $\psi(2S)$ are included).

and therefore the upper mass sideband will not be a good representation of the background line shape and the estimation of combinatorial background in the signal region could be inaccurate. For this reason, no veto is applied and instead a component for these backgrounds is included in the invariant mass fit model, further discussed in Sec. 7.1.3.

5.3 Neural network binning

When making use of a multivariate classifier to separate signal and background in an analysis, there are two main techniques in optimising the output of the classifier. The first is to apply selection criteria on the output where the value of the criteria are optimised for some figure of merit, typically a measure of the expected significance of the searched for signal. The second is to instead use the classifier output in the analysis itself, *e.g.* by performing a fit to the output or by performing a fit to the invariant mass of the candidates in bins of the output, where the binning scheme itself is optimised rather than selection criteria. The advantage of this second technique is that it preserves additional signal events that would otherwise be discarded using the first technique. In the case of performing the fit in bins of classifier output, information from the purer bins of output allow a more stable fit to the less pure bins to be performed. As the decay mode being searched for is very rare, in this analysis a fit in bins of classifier output is performed.

At this point, it is useful to redefine the classifier output as determined in Sec. 5.1.4 in terms of its efficiency on the signal mode. The output is flattened such that a result from 0 to 1 is equivalent to one minus the efficiency on the signal (*i.e.* 0.4 is equivalent to a signal efficiency of 60%). As discussed in Sec. 5.1.4, an initial loose cut on the neural network output is applied that is highly efficient on the signal which corresponds to a cut at 0.05 in this flattened output. The remaining candidates after this cut are then divided into bins of classifier output, with equal amounts of signal (as determined from simulation) in each bin. To determine the optimal number of bins a number of pseudo experiments are performed, generating events based on probability density functions (PDFs) and yields for a number of different components. Four components are included in these toy experiments: a component for $B^0 \to K^{*0}\mu^+\mu^-$ decays where the yield is taken from a blind fit to the rare mode; a component for the $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ signal with the yield set as 1/100 of the $B^0 \to K^{*0}\mu^+\mu^-$ yield (based on a simple scaling due to the CKM ratio $|V_{td}/V_{ts}|^2$ and the B_s/B^0 production ratio); a component for the combinatorial background where the yield is taken from the blind fit to the rare mode; and a component for the background from $\Lambda_b^0 \to pK^-\mu^+\mu^-$, where the yield is extracted from a data sample (the exact method of which is discussed later in Sec. 7.1.4). Six binning schemes are then tested, with 5,000 pseudo experiments being performed for each one and the significance of the fitted signal component being determined each time by Wilks' theorem [115]. The average significance over all performed pseudo experiments is then used as the figure of merit to choose the optimal binning scheme.

The results of the toy studies carried out over all binning schemes are shown in Tab. 5.5. From the tested binning schemes the highest significances are those with four, five or six bins, with all having consistent figures of merit. To simplify the fitting procedure, the least number of bins is taken and so the number of bins used is four. Also included is a result of optimising the classifier output by the first method described above, choosing an optimal cut. The optimal binning scheme has a significance approximately 7% higher then the optimised cut, validating the choice to bin in classifier output.

The studies to determine an optimal binning scheme are carried out using only the Run 1 sample size. This is due to the neural network classifier being different between Run 1 and Run 2 and thus not allowing the datasets to be combined and binned at the same time. Instead, the optimal number of four bins is found for the single Run 1 dataset and the second dataset will be included as an additional four bins, leading to a total of eight bins in the fit to invariant mass. The boundaries of the optimised binning scheme are [0.05, 0.2875], [0.2875,0.525], [0.525, 0.7625] and [0.7625, 1.00]. Figure 5.22 shows the classifier output on simulated signal samples as well as the distribution of background events and the location of the bin boundaries for both Run 1 and Run 2.

One of the key assumptions in this binning scheme is that the efficiency in bins of neural network output is the same between $B^0 \to K^{*0}\mu^+\mu^-$ and $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$. This assumption allows the more precise parameters from the much larger number of $B^0 \to K^{*0}\mu^+\mu^-$ decays to fix those from $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$. To test this assumption, the efficiencies in bins of classifier are validated on simulation. Figure 5.23 shows these efficiencies for the two decays modes plus their equivalent control modes for both Run 1 and Run 2. It is clear that, within statistical

Table 5.5: Significance of observing 0.01 $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ events for every $B^0 \to K^{*0}\mu^+\mu^-$ event for all tested binning schemes.

Binning Scheme	Significance
1 bin with optimised cut	2.29
2 bins	2.28
3 bins	2.39
4 bins	2.45
5 bins	2.46
6 bins	2.45
7 bins	2.31

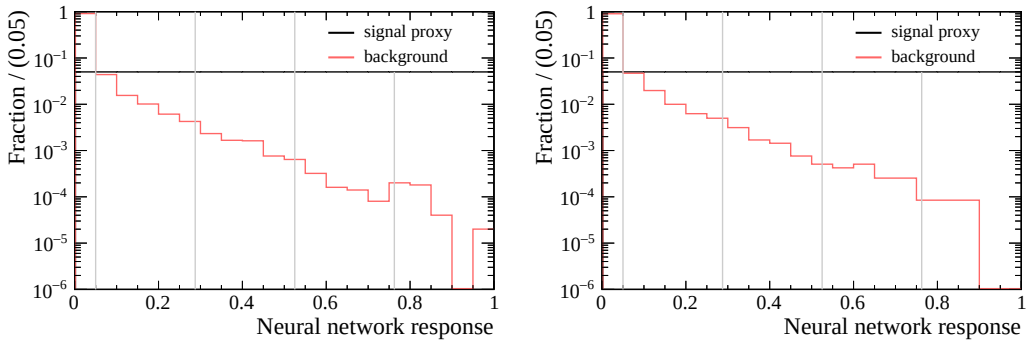

Figure 5.22: Neural network responses after flattening of the network with respect to the signal output for both simulated signal decays and combinatorial background events for Run 1 (left) and Run 2 (right) data. Binning boundaries are indicated with vertical lines and events below 0.05 are rejected in the final analysis.

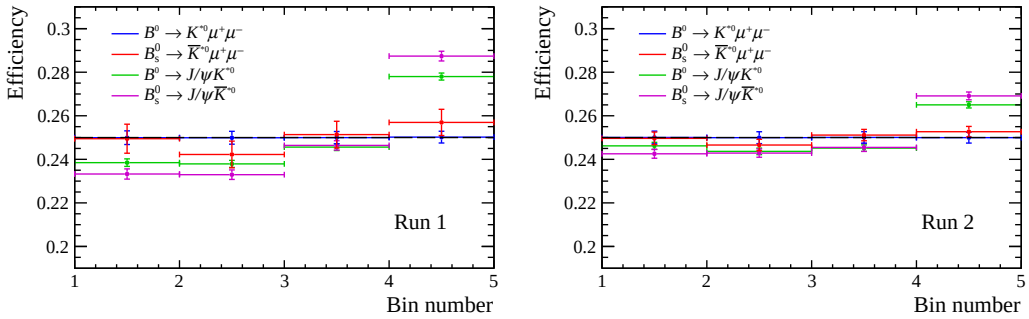

Figure 5.23: Efficiencies in bins of neural network output for the rare and control decay modes for Run 1 (left) and Run 2 (right). The black horizontal line indicates equal efficiency across all bins and the legend details the contributions from each decay mode.

uncertainties, the binned efficiencies are in agreement between B^0 and B_s modes. There is a noticeable difference between the rare modes and the control modes, this is expected due the difference in kinematics because of the presence of the J/ψ resonance.

5.4 Particle identification optimisation

For the purposes of carrying out background studies and dealing with a more manageably sized dataset, initial PID requirements have been applied at $\texttt{ProbNN}_\pi \times (1 - \texttt{ProbNN}_K) > 0.05$ for the pion and $\texttt{ProbNN}_K \times (1 - \texttt{ProbNN}_\pi) > 0.05$ for the kaon. These cuts reduce the most common particle misidentification, that of a kaon to a pion and vice-versa. The background studies that have been performed have shown that there are still remaining peaking backgrounds in the dataset, in particular that

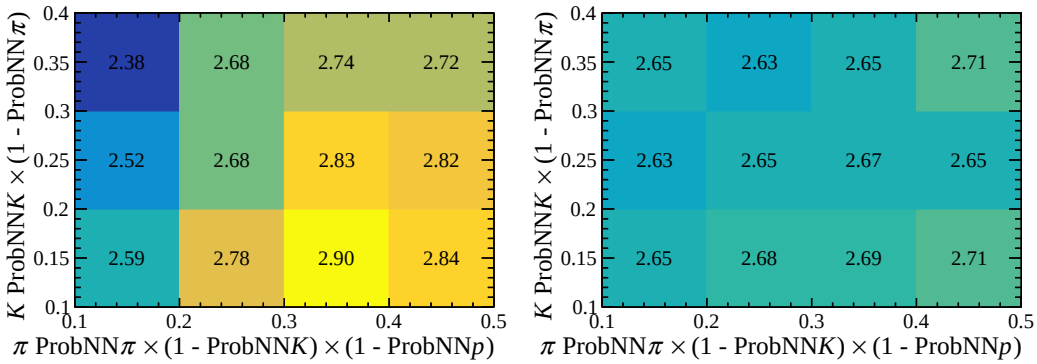

Figure 5.24: Significance of observing 0.01 $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ decays for every $B^0 \to K^{*0}\mu^+\mu^-$ decay, taken as an average from 2000 pseudo-experiments for Run 1 (left) and Run 2 (right).

of $\Lambda_b^0 \to pK^-\mu^+\mu^-$ which cannot be removed by a veto. The most effective way to reduce this background is the make use of tighter requirements on PID variables and to explicitly target proton to pion misidentification. As such, the combined PID variable that will be tested is $\texttt{ProbNN}_\pi \times (1 - \texttt{ProbNN}_K) \times (1 - \texttt{ProbNN}_p)$ for the pion while keeping the above combination for the kaon.

To perform an optimisation of requirements on PID variables, the same method of using pseudo-experiments detailed above for choosing an optimal neural network output binning scheme is used. The same components are used in this model, with the component for $\Lambda_b^0 \to pK^-\mu^+\mu^-$ being key to the optimisation. The yields of the components taken from the above studies are scaled by using efficiencies taken from PIDCalib, with efficiencies for each component calculated at each investigated working point. While this method is effective for optimising the PID to reduce the impact of known background components, it is blind to the effect of any component not included in the model. It is not expected that there will be any large contributions from other background, but to reduce any potential impact the minimum combined PID requirements are set at greater than 0.1.

The results of these pseudo-experiments are shown in Fig. 5.24. For the Run 1 study there is a clear optimal value at $\texttt{ProbNN}_K \times (1 - \texttt{ProbNN}_\pi) > 0.1$ for the kaon and $\texttt{ProbNN}_\pi \times (1 - \texttt{ProbNN}_K) \times (1 - \texttt{ProbNN}_p) > 0.3$ for the pion. The Run 2 study shows no optimal value, with the figure of merit being essentially the same for each working point. This is not unexpected as the tuning of the neural network based PID variables are different between the Run 1 and Run 2 dataset. The performance of the Run 2 variables has been found to be better then those for Run 1 and so not as much is gained from tighter PID requirements (the looser requirements already rejecting most of the background events from $\Lambda_b^0 \to pK^-\mu^+\mu^-$). As the significances

are found to be consistent between the investigated values, the same value for PID selection is used between Run 1 and Run 2 for simplicity.

5.5 Selected candidates

The reconstructed invariant mass of all of the candidates that have passed the selection criteria detailed in this chapter are shown in Fig. 5.25 for the rare mode and Fig. 5.26 for the control mode, where the combination of all eight bins of neural network response over both Run 1 and Run 2 is shown. For the control mode candidates, the two invariant mass reconstructions are shown, both with and without constraining the J/ψ mass.

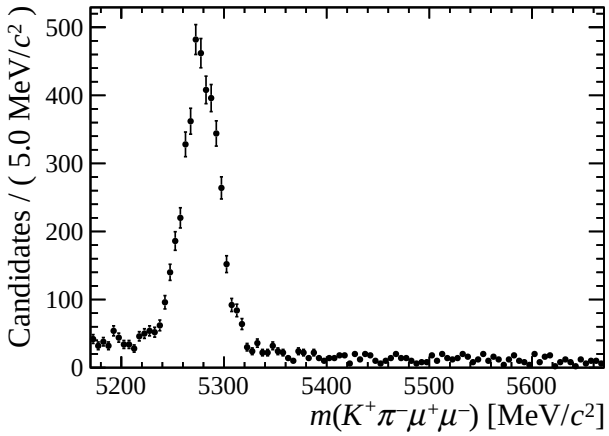

Figure 5.25: Invariant mass distribution of candidates which have passed the selection criteria in the rare mode.

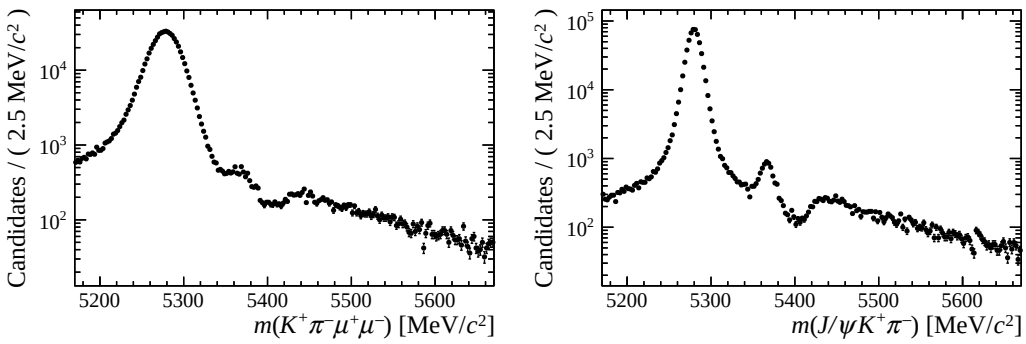

Figure 5.26: Invariant mass distribution of candidates which have passed the selection criteria in the control mode, both without (left) and with (right) a constraint on the J/ψ mass applied. Results are shown on a logarithmic scale.

Clearly, in all of the datasets, the contribution from B^0 decays is dominant. In the rare mode, the combinatorial background component is still clearly present and leads to there being no visible B_s^0 component. In the control modes, the candidates are shown on a logarithmic scale to highlight components other than the B^0. A peaking structure is visible from B_s^0 decays and in the upper mass sideband there is a clear raised structure above that from the combinatorial background. This structure is a result of the $B^+ \to J/\psi K^+$ decays which have not been vetoed in this dataset. It is clear from these plots that the application of the J/ψ mass constraint greatly improves the invariant mass resolution and allows the B^0 and B_s^0 to be more clearly separated. It can also be seen that the mass constraint affects the invariant mass shape of $B^+ \to J/\psi K^+$ decays, causing them to be more clearly separated from the combinatorial background at the threshold of the decay mode. The differences in line-shapes of decays with and without the J/ψ mass constraint will be further considered in Sec. 7.1.

Chapter 6

Efficiencies

To measure the branching fraction ratio between two modes, it is essential that the efficiency to select candidates from each decay is measured correctly. The following chapter presents the total selection efficiencies for the decay modes $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$, $B^0 \to J/\psi K^{*0}$, $B^0 \to K^{*0}\mu^+\mu^-$ and $B_s^0 \to J/\psi\overline{K}^{*0}$ as well as the details of how the individual components of the total efficiency are calculated.

The efficiency for each decay mode is calculated by making use of MC simulated samples and is calculated separately for each year and magnet polarity in the data sample. A weighted average of these is then used in the branching fraction calculation, taking into account the relative sizes of each dataset.

It should be noted here that the efficiencies for the $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ mode 2011 data are not taken from MC due to a lack of available simulated samples. Instead, these efficiencies are calculated by taking the relative efficiencies of the $B^0 \to K^{*0}\mu^+\mu^-$ mode between 2011 and 2012 and then multiplying the 2012 $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ efficiency. This is done under the assumption that the relative efficiencies between the year is the same for these modes. This is validated by checking that this also true for the control modes $B^0 \to J/\psi K^{*0}$ and $B_s^0 \to J/\psi\overline{K}^{*0}$.

The efficiencies are calculated from the product of four separate contributions: the geometrical efficiency for the particle decay to occur within the LHCb acceptance; the stripping and reconstruction efficiency which accounts for the efficiency of the selection criteria within the stripping line as well as the efficiency to find and reconstruct the four final state tracks; the trigger efficiency of the requirements within the trigger lines used in this analysis; and finally the selection efficiency, accounting for the particle identification, neural network output and vetoes requirements applied to the candidates. The efficiencies are calculated in a sequential manner, with the efficiency of each contribution being calculated after

the application of the previous contribution.

The total efficiency is calculated as

$$\epsilon^{\text{total}} = \epsilon^{\text{geom}} \times \epsilon^{\text{reco\&strip|geom}} \times \epsilon^{\text{trigger|reco\&strip\&geom}} \times \epsilon^{\text{sel|trigger\&reco\&strip\&geom}}. \quad (6.1)$$

The full list of these efficiencies, divided by decay mode, year and polarity as well as the combined total efficiencies are shown in Tab. 6.2.

6.1 Geometrical and stripping efficiencies

The geometrical efficiency of each decay mode accounts for the efficiency of the modes to decay within and all the final state particles to be within the LHCb detector acceptance as given in Sec. 3.2. This is calculated at the time that the MC samples are generated using the known information on the detector acceptance and only simulating decays with final state particles in the acceptance. The stripping and reconstruction efficiencies are also calculated using simulation, with the ratio of yields in each mode from before and after passing the stripping line requirements and being reconstructed taken as the efficiency.

6.2 Trigger efficiencies

The trigger efficiencies are once again calculated from simulated samples using the yield of candidates before and after the application of the trigger line requirements. In addition to this calculation taken from simulation, this efficiency can be taken directly from the data as a validation of the simulated efficiency. The `L0Muon` efficiency can be determined from data using $B^0 \to J/\psi K^{*0}$ candidate and the so called TISTOS method which takes advantage of the fact that an event can be selected independently of the signal candidate. The `L0Muon` efficiency is computed according to

$$\epsilon_{\text{TOS}} = \frac{N_{\text{TIS\&TOS}}}{N_{\text{TIS}}} \quad (6.2)$$

where N_{TIS} is the number of candidates that have been triggered by particles in the event that have not been used to form the candidate and $N_{\text{TIS\&TOS}}$ is the subset of these candidates that have also been triggered by `L0Muon`. This efficiency is computed under the assumption that the TIS and TOS decisions are uncorrelated. The TIS requirements are that either `L0Muon` or `L0Hadron` has been triggered independently of the candidate. It is also required that the SPD multiplicity is less

than 600 or 450 for Run 1 and Run 2 respectively to ensure the multiplicity requirements align with those used in `L0Muon`. This method is carried out both on data and on simulation, with the ratio between the two measurements being used as the validation. The results of this are shown in Fig. 6.1.

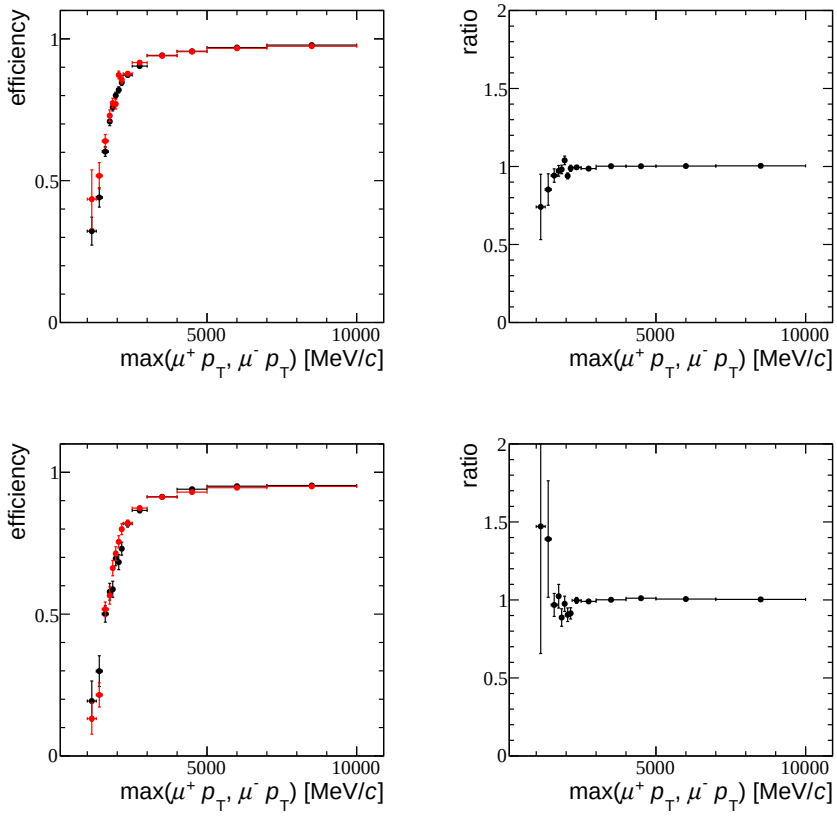

Figure 6.1: Trigger efficiency as a function of maximum p_{T} of the μ^+ or μ^- for $B^0 \to J/\psi K^{*0}$ candidates in 2011+2012 (top) and 2016 (bottom). Black points show data with efficiency calculated via the TISTOS method while the red shows the TISTOS efficiency on the MC. The right-hand figures show the data/MC ratio.

The weights derived in the above procedure are then applied to simulation and the efficiency of the `L0Muon` trigger is calculated both with and without the weights. For the Run 1 dataset, the efficiencies with/without the weights applied are consistent with each other (0.7840 ± 0.0028 and 0.7877 ± 0.0028) and as such no correction is applied. In the Run 2 dataset there is a larger discrepancy between data and simulation (0.798 ± 0.004 and 0.769 ± 0.004 with and without the applied weights) and as such the efficiency with corrections applied are used. The primary reason for the data-simulation differences being larger in Run 2 is because the trigger

conditions were more variable in the 2016 data taking period. The simulation is generated according to one specific set of trigger conditions and therefore does not necessarily match the average of the data.

6.3 Selection efficiencies

The selection efficiency for each mode is further divided into three parts: the efficiency of the vetoes and non-PID selection requirements described in Sec. 5.1, the efficiency of the PID selection requirements and the efficiency on the modes due to the mass window over which the invariant mass fit is carried out. The values of these efficiencies are detailed in Tab. 6.1.

Whereas the veto efficiencies can be evaluated directly from the simulated samples in the same manner described for the above efficiencies, the same cannot be done for the PID efficiencies. Particle identification response is a known area where the simulated response does not match with the true response on data. As such, the PidCalib package (introduced and briefly described in Sec. 5.2.1) is used instead. By taking samples of decays where the species particles can be cleanly identified independently of the PID systems, that is the kinematics of the decay make them easily identifiable, the efficiency of the PID variables can be tested. The efficiency of a given requirement can be evaluated in the bins of kinematic variables for the simulated samples and then averaged to give an overall efficiency.

There are two uncertainties associated with using PidCalib to evaluate these efficiencies. The first uncertainty is a statistical uncertainty related to the size of the calibration sample available for a given particle species. The second is a systematic uncertainty related to the binning scheme chosen when evaluating the efficiencies. As the response is measured in bins of kinematic variables, the granularity of the binning scheme and the location of the bin boundaries can have a significant affect on the final result. To test the size of this effect, an alternative binning scheme is used where one less bin is used in each of the kinematic variables (track pseudo-rapidity, track momenta and either track multiplicity or SPD multiplicity depending on whether it is for Run 1 or Run 2 data). The difference between the efficiency calculated with the standard binning scheme and the alternate binning scheme is then taken as the systematic uncertainty on the measurement. The size of this uncertainty can be up to 8% on the total PID efficiency for a given mode, however at no point is a single mode used in the normalisation calculation of the rare mode. Instead ratios of efficiencies are used and therefore the correlation between the binning scheme uncertainties of the two decay modes needs to be taken into account. This greatly

reduces the systematic uncertainty.

The final part of the selection efficiency is that related to the invariant mass window over which the fit will be carried out, in this analysis $5170\,\mathrm{MeV}/c^2 < m(K^+\pi^-\mu^+\mu^-) < 5670\,\mathrm{MeV}/c^2$. This efficiency needs to be considered separately depending on whether a fit is carried out with or without a constraint on the J/ψ mass in the control modes, as this affects the resolution of the signal and thus the relative amount of the signal in the fit window. The values of these efficiencies as determined from simulation are shown in Tab. 6.1. For the total efficiency used in the branching fraction determination, the values with a J/ψ mass constraint are used for the control mode. The values without a mass constraint are used to determine a systematic uncertainty associated to the line shape of the signal, further discussed in Chapter 8.

6.4 Corrections to efficiencies

As previously detailed in Sec. 5.1.3, corrections to the B meson p_T, B meson vertex χ^2 and detector occupancy have been derived from data for the simulated samples. These same corrections are applied when calculating the efficiencies, as well as the corrections related to the Level 0 trigger detailed above. In addition, two more sets of corrections are taken into account: one for the track reconstruction efficiency difference between data and simulation and the other for the PID requirements applied in the stripping line selection (`IsMuon` and `PIDmu` > -3). These corrections are applied in the same way as the previous ones, as weights where the weight corresponds to the relative efficiency between data and MC.

For the track reconstruction efficiency, a correction for the long track efficiency is derived centrally by the LHCb collaboration. These corrections are determined through the same tag-and-probe approach with J/ψ decays described in Sec. 3.2.5. The values of this correction in bins of track momenta and pseudorapidity are shown in Fig. 6.2 These are applied to each of the final state tracks after reconstruction/stripping. The corrections have been derived from muon tracks and in principle the uncertainties on hadronic tracks, in this case the final state kaon and pion, will not be the same as for the muons. However, as at all points in this analysis ratios of efficiencies are taken, the differing uncertainties are reduced to a negligible level and the corrections derived from muon tracks can be used.

For the muon identification requirements, the efficiency in simulation is determined with the same selection criteria applied as used for the PidCalib samples. The ratio of MC and PidCalib efficiencies is then computed and used to weight

each of the muons in the simulated samples. The efficiency of the stripping is then calculated with these weights applied.

The effect of both the tracking and PID correction weights almost completely cancels when taking the ratio between different decays. The remaining effect is included as a source of systematic uncertainty, discussed further in Chapter 8.

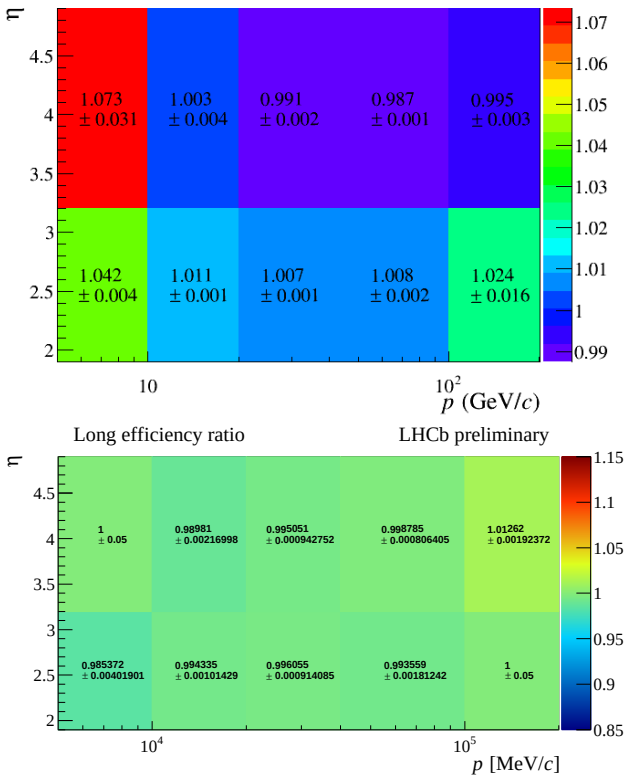

Figure 6.2: Tracking efficiency corrections in bins of track momenta and pseudo-rapidity for Run 1 (top) and Run 2 (bottom) determined from data and simulation using a tag-and-probe approach.

Table 6.1: Veto, PID and invariant mass fit window efficiencies for the different decay modes used in this analysis, split by year and polarity.

Decay mode	Year	Polarity	ϵ^{veto}	ϵ^{PID}	$\epsilon^{\text{fitwindow}}$	
					J/ψ constraint	no J/ψ constraint
$B^0 \to K^{*0}\mu^+\mu^-$	2016	Down	0.466 ± 0.004	0.83 ± 0.03	-	0.976 ± 0.011
		Up	0.471 ± 0.004	0.83 ± 0.03	-	0.975 ± 0.010
	2012	Down	0.512 ± 0.007	0.78 ± 0.04	-	0.968 ± 0.015
		Up	0.510 ± 0.007	0.77 ± 0.04	-	0.966 ± 0.015
	2011	Down	0.524 ± 0.005	0.77 ± 0.04	-	0.973 ± 0.010
		Up	0.523 ± 0.005	0.77 ± 0.04	-	0.971 ± 0.010
$B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$	2016	Down	0.470 ± 0.004	0.83 ± 0.03	-	0.987 ± 0.011
		Up	0.470 ± 0.004	0.84 ± 0.03	-	0.988 ± 0.010
	2012	Down	0.510 ± 0.007	0.78 ± 0.04	-	0.985 ± 0.015
		Up	0.518 ± 0.007	0.78 ± 0.05	-	0.986 ± 0.015
$B^0 \to J/\psi K^{*0}$	2016	Down	0.762 ± 0.004	0.83 ± 0.03	0.977 ± 0.006	0.976 ± 0.006
		Up	0.764 ± 0.004	0.84 ± 0.03	0.976 ± 0.006	0.975 ± 0.006
	2012	Down	0.816 ± 0.007	0.79 ± 0.06	0.974 ± 0.008	0.977 ± 0.008
		Up	0.817 ± 0.007	0.78 ± 0.06	0.973 ± 0.008	0.975 ± 0.008
	2011	Down	0.822 ± 0.006	0.78 ± 0.06	0.977 ± 0.006	0.976 ± 0.006
		Up	0.825 ± 0.006	0.78 ± 0.05	0.977 ± 0.006	0.976 ± 0.006
$B_s^0 \to J/\psi \overline{K}^{*0}$	2016	Down	0.761 ± 0.005	0.83 ± 0.03	0.979 ± 0.007	0.992 ± 0.008
		Up	0.765 ± 0.006	0.84 ± 0.03	0.980 ± 0.008	0.991 ± 0.008
	2012	Down	0.814 ± 0.009	0.80 ± 0.06	0.975 ± 0.012	0.991 ± 0.012
		Up	0.816 ± 0.009	0.79 ± 0.06	0.975 ± 0.011	0.990 ± 0.011
	2011	Down	0.821 ± 0.009	0.79 ± 0.06	0.975 ± 0.011	0.993 ± 0.011
		Up	0.826 ± 0.009	0.79 ± 0.06	0.978 ± 0.011	0.992 ± 0.011

Table 6.2: Efficiencies divided into individual components as well as the total efficiency for all decay modes, given for each year and polarity.

Decay mode	Year	Polarity	ϵ^{geom}	$\epsilon^{\text{reco\&strip}}$	$\epsilon^{\text{trigger}}$	ϵ^{sel}	ϵ^{total}
$B^0 \to K^{*0}\mu^+\mu^-$	2016	Down	0.1699 ± 0.0005	0.0974 ± 0.0005	0.755 ± 0.006	0.376 ± 0.016	$(4.70 \pm 0.21) \times 10^{-3}$
		Up	0.1712 ± 0.0006	0.0966 ± 0.0005	0.761 ± 0.005	0.383 ± 0.016	$(4.82 \pm 0.21) \times 10^{-3}$
	2012	Down	0.1646 ± 0.0012	0.0831 ± 0.0006	0.759 ± 0.008	0.384 ± 0.021	$(3.99 \pm 0.23) \times 10^{-3}$
		Up	0.1645 ± 0.0009	0.0831 ± 0.0006	0.758 ± 0.008	0.379 ± 0.023	$(3.93 \pm 0.25) \times 10^{-3}$
	2011	Down	0.1606 ± 0.0008	0.0931 ± 0.0005	0.736 ± 0.005	0.392 ± 0.023	$(4.31 \pm 0.26) \times 10^{-3}$
		Up	0.1599 ± 0.0009	0.0934 ± 0.0005	0.742 ± 0.005	0.389 ± 0.022	$(4.31 \pm 0.25) \times 10^{-3}$
$B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$	2016	Down	0.1691 ± 0.0005	0.0985 ± 0.0005	0.787 ± 0.006	0.385 ± 0.015	$(5.05 \pm 0.21) \times 10^{-3}$
		Up	0.1694 ± 0.0006	0.0983 ± 0.0005	0.787 ± 0.006	0.388 ± 0.015	$(5.09 \pm 0.20) \times 10^{-3}$
	2012	Down	0.1638 ± 0.0006	0.0851 ± 0.0006	0.769 ± 0.008	0.392 ± 0.024	$(4.20 \pm 0.26) \times 10^{-3}$
		Up	0.1636 ± 0.0006	0.0844 ± 0.0006	0.777 ± 0.008	0.398 ± 0.027	$(4.26 \pm 0.29) \times 10^{-3}$
$B^0 \to J/\psi K^{*0}$	2016	Down	0.1679 ± 0.0005	0.1024 ± 0.0004	0.809 ± 0.004	0.620 ± 0.024	$(8.62 \pm 0.33) \times 10^{-3}$
		Up	0.1664 ± 0.0006	0.1029 ± 0.0004	0.808 ± 0.004	0.626 ± 0.023	$(8.66 \pm 0.33) \times 10^{-3}$
	2012	Down	0.1611 ± 0.0006	0.0875 ± 0.0004	0.784 ± 0.006	0.629 ± 0.046	$(6.95 \pm 0.51) \times 10^{-3}$
		Up	0.1605 ± 0.0006	0.0868 ± 0.0004	0.782 ± 0.006	0.624 ± 0.045	$(6.80 \pm 0.49) \times 10^{-3}$
	2011	Down	0.1584 ± 0.0007	0.0987 ± 0.0005	0.764 ± 0.005	0.628 ± 0.046	$(7.50 \pm 0.55) \times 10^{-3}$
		Up	0.1572 ± 0.0007	0.0995 ± 0.0005	0.763 ± 0.005	0.627 ± 0.042	$(7.48 \pm 0.51) \times 10^{-3}$
$B_s^0 \to J/\psi \overline{K}^{*0}$	2016	Down	0.1675 ± 0.0009	0.1094 ± 0.0005	0.815 ± 0.005	0.620 ± 0.025	$(9.26 \pm 0.38) \times 10^{-3}$
		Up	0.1676 ± 0.0006	0.1093 ± 0.0005	0.820 ± 0.005	0.628 ± 0.025	$(9.44 \pm 0.38) \times 10^{-3}$
	2012	Down	0.1615 ± 0.0009	0.0922 ± 0.0007	0.776 ± 0.008	0.632 ± 0.046	$(7.30 \pm 0.54) \times 10^{-3}$
		Up	0.1617 ± 0.0006	0.0914 ± 0.0006	0.777 ± 0.008	0.627 ± 0.046	$(7.21 \pm 0.54) \times 10^{-3}$
	2011	Down	0.1577 ± 0.0006	0.1040 ± 0.0007	0.769 ± 0.007	0.632 ± 0.046	$(7.96 \pm 0.59) \times 10^{-3}$
		Up	0.1583 ± 0.0006	0.1054 ± 0.0007	0.765 ± 0.007	0.635 ± 0.046	$(8.11 \pm 0.59) \times 10^{-3}$

Chapter 7

Invariant mass fit

The yield of $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ decays is determined by performing an invariant mass fit to the $m(K^+\pi^-\mu^+\mu^-)$ spectrum. In addition, to make a measurement of the branching fraction of this mode a fit to the $m(J\psi K^+\pi^-)$ spectrum needs to be performed to extract the yields of the control mode. An invariant mass fit of this kind requires individual PDFs to describe each of the components of the fit. The following chapter describes how PDFs for the fit components are determined. This is then followed by a description of how these components are combined to carry out the full fit to the invariant mass spectrum and the results of this fit. Finally, the details and results of the branching fraction calculation are given.

7.1 Fit components

7.1.1 Signal component

In this analysis, both decays of B_s^0 and B^0 mesons going to a $K^+\pi^-\mu^+\mu^-$ final state have similar mass distributions and can only be distinguished by the difference in the mass of the parent particle. This is particularly useful here because it allows the $B^0 \to K^{*0}\mu^+\mu^-$ mode to be used to fix the parameters of any component describing the $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ mode and the same PDF can be used to describe both. As such, the term signal component is used here to refer to components describing both the $B^0 \to K^{*0}\mu^+\mu^-$ ($B^0 \to J/\psi K^{*0}$) and the $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ ($B_s^0 \to J/\psi \overline{K}^{*0}$) in the fit to the rare mode (control mode).

To describe these modes, a PDF made up of the sum of two Crystal-Ball (CB) functions [130] and a Gaussian function are used, sharing a common mean between all three components. The two Crystal-Ball functions have power law tails on opposite sides in terms of invariant mass. The first tail is on the lower side

and is present to describe the effects of the final state particles (specifically the muons) radiating extra energy which is lost in the mass reconstruction. The second tail on the upper side describes candidates which are not as well reconstructed at high invariant mass and thus have a non-Gaussian distribution. The inclusion of the additional Gaussian component accounts for the fact that the resolution of the reconstructed mass is not constant over all the candidates and allows for a variation in the per-event resolution. The exact functional form of the PDF for the B^0 component, where F_{CB} and F_{Gauss} represent the constituent CB and Gaussian components respectively, is given by:

$$\begin{aligned} S_{B^0}(m) = & f_{DCB} \cdot [f_{CB1} \cdot F_{CB}(m; \mu_{B^0}, \sigma_{CB}, \alpha_{CB1}, n_{CB1}) + \\ & (1 - f_{CB1}) \cdot F_{CB}(m; \mu_{B^0}, \sigma_{CB}, \alpha_{CB2}, n_{CB2})] + \\ & (1 - f_{DCB}) \cdot F_{Gauss}(m; \mu_{B^0}, \sigma_{Gauss}) \end{aligned} \tag{7.1}$$

where μ_{B^0} is the common mean of the component PDFs, σ_{CB} is the shared width of the CB functions, σ_{Gauss} is the width of the Gaussian component and $\alpha_{CB(i)}$ and $n_{CB(i)}$ are the tail parameters of the two CB functions. The form of the PDF to describe B^0_s candidates ($S_{B^0_s}$) differs from S_{B^0} only by its mean, $\mu_{B^0_s}$, which is fixed w.r.t. μ_{B^0} according to the measurement made in Ref. [94]:

$$\mu_{B^0_s} = \mu_{B^0} + 87.5\,\mathrm{MeV}/c^2$$

As the invariant mass fit is carried out in bins of neural network response and there can be some potential variation in the line-shape of the signal in these bins, in particular in the shapes of the CB tails, individual PDFs are used to describe the signal in each neural network bin. The parameters that will be used in the fit to the rare mode are determined in a two step procedure. First, individual fits in each bin of neural network response are carried out on simulated samples of $B^0 \to J/\psi K^{*0}$ candidates to fix the relative fractions in each of the component PDFs, f_{CB} and f_{DCB}, the relative width between the CB components and the Gaussian component, $\sigma_{Gauss}/\sigma_{CB}$, and one of the CB tail parameters for each CB function, α_{CB1} and α_{CB2}. The tail parameters n_{CB1} and n_{CB2} are fixed to a common value over all response bins when performing these fits, determined from a fit to the combination of all the bins. Using these fixed parameters, a fit the $B^0 \to J/\psi K^{*0}$ mode data is carried out (with the rest of the components in this fit being described in the rest of this section) and the mean of the PDF, width of the CB functions, and a scaling of tail parameters over all bins of neural networking response, α^{scale}_{CB1} and α^{scale}_{CB2}, are determined. This second fit is carried out without a constraint on the J/ψ mass

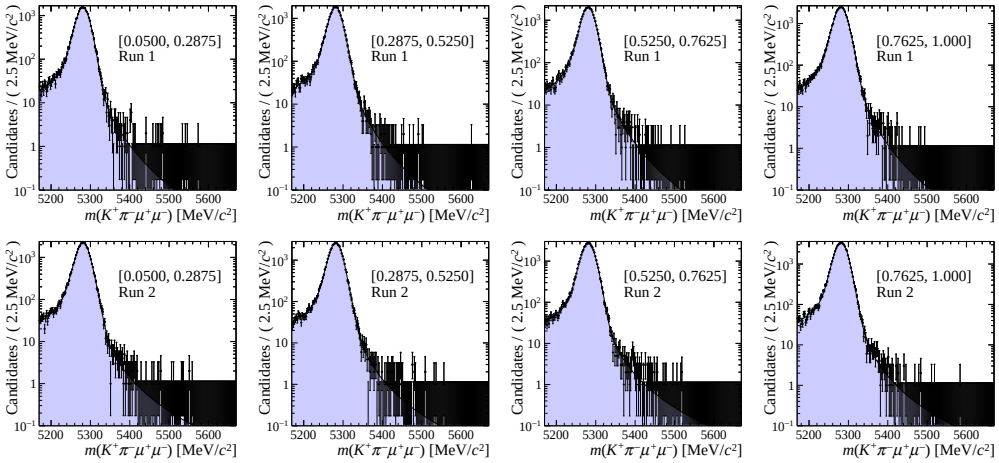

Figure 7.1: Fit in bins of increasing neural network response to $B^0 \to J/\psi K^{*0}$ MC for Run 1 (top four) and Run 2 (bottom four). Results are shown on a log scale.

applied, allowing the $B^0 \to J/\psi K^{*0}$ mode to act as a proxy for the $B^0 \to K^{*0}\mu^+\mu^-$ mode.

The motivation for why some of the parameters of the signal PDFs are determined directly from simulation and why some are scaled by the result of the fit to the control mode is mainly due to known differences in the modelling of the detector resolution in the simulation. The generic line-shape is fixed from simulation, but a scaling of the resolution can be determined from data to ensure that the resolution of the signal component is correct in the rare mode where sensitivity is paramount. There is also the potential for the simulated samples to not correctly describe the tail parameters of the CB components. By determining an initial value of the tail parameters α_{CB1} and α_{CB2} from simulation and then allowing a scaling from the fit to the control mode, any small difference between simulation and data can be accounted for.

The results of the fit to the simulated samples are shown in Fig. 7.1, with the resulting parameters from these fits shown in Tab. 7.1. It is clear that while the general shape of the components are similar, there is some variation between bins as well as a variation depending on whether it is Run 1 or Run 2 data being considered. By using separate components in each bin the signal component can be well modelled.

In addition to the fit to the $B^0 \to J/\psi K^{*0}$ mode to determine the fit parameters for the rare mode, it is also necessary to fit the $B^0 \to J/\psi K^{*0}$ mode to extract a yield that will be used in the normalisation procedure. For this purpose it is much more beneficial to carry out the fit with a constraint on the J/ψ mass

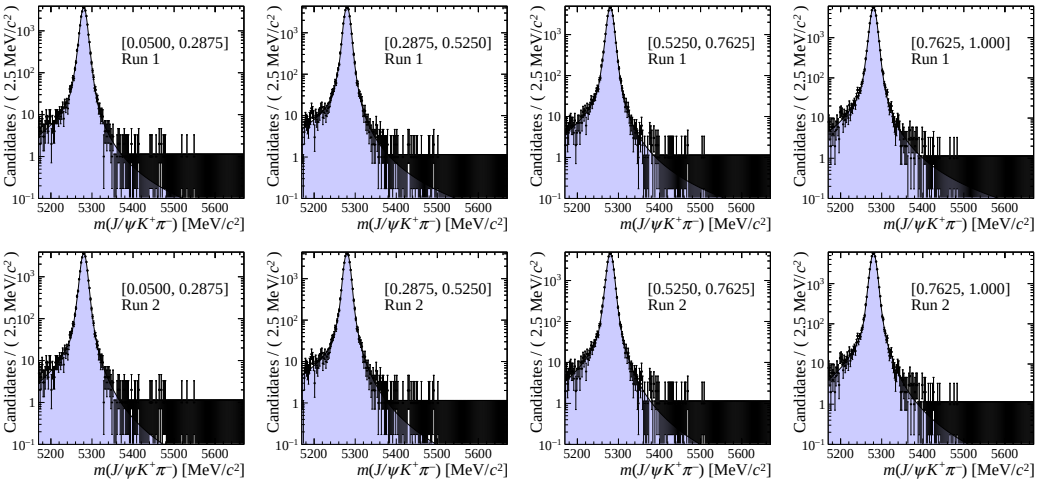

Figure 7.2: Fit in bins of increasing neural network response to $B^0 \to J/\psi K^{*0}$ MC with a constraint on the J/ψ mass applied for Run 1 (top four) and Run 2 (bottom four). Results are shown on a log scale.

applied, greatly improving the resolution of the signal components and thus the statistical uncertainty on the B^0 and B_s^0 yields. The same general PDF is used for this fit, the sum of two CB functions and a Gaussian, however the parameters will be significantly different and need to be independently determined from simulation. As before, some initial parameters are extracted from simulation with a scaling of the width and tail parameters allowed at a later stage to account for differences between data and MC. The results of the fits to simulation are shown in Fig. 7.2 with the fit parameters shown in Tab. 7.1.

Table 7.1: Parameters of double Crystal Ball and Gaussian shape obtained from fit to $B^0 \to J/\psi K^{*0}$ MC in bins of neural network response. Labels 1 and 2 distinguish between the two CB components. The n parameter for each tail has been fixed to the shown value.

Parameter	Run 1				Run 2			
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 1	Bin 2	Bin 3	Bin 4
	No mass constraint							
μ	5280.84 ± 0.14	5280.76 ± 0.13	5280.64 ± 0.11	5280.70 ± 0.10	5281.20 ± 0.11	5281.26 ± 0.10	5280.97 ± 0.10	5281.08 ± 0.08
σ_{CB}	12.5 ± 0.7	11.8 ± 0.7	13.9 ± 0.4	13.0 ± 0.5	13.2 ± 0.4	13.5 ± 0.4	13.2 ± 0.5	13.81 ± 0.32
α_{CB1}	-0.99 ± 0.17	-1.07 ± 0.13	-1.17 ± 0.15	-1.31 ± 0.12	-1.51 ± 0.14	-1.82 ± 0.10	-1.49 ± 0.14	-1.80 ± 0.11
α_{CB2}	1.07 ± 0.11	1.02 ± 0.14	1.42 ± 0.10	1.17 ± 0.12	1.15 ± 0.11	1.09 ± 0.12	1.07 ± 0.12	1.18 ± 0.10
n_{CB1}	4.5	4.5	4.5	4.5	3.2	3.2	3.2	3.2
n_{CB2}	1.6	1.6	1.6	1.6	1.5	1.5	1.5	1.5
$\sigma_{Gauss}/\sigma_{CB}$	1.35 ± 0.07	1.45 ± 0.07	1.29 ± 0.06	1.35 ± 0.04	1.45 ± 0.04	1.40 ± 0.05	1.30 ± 0.05	1.33 ± 0.04
f_{CB1}	0.24 ± 0.08	0.33 ± 0.07	0.20 ± 0.07	0.36 ± 0.08	0.34 ± 0.09	0.53 ± 0.08	0.38 ± 0.08	0.50 ± 0.07
f_{DCB}	0.50 ± 0.07	0.53 ± 0.08	0.75 ± 0.08	0.60 ± 0.07	0.63 ± 0.06	0.73 ± 0.07	0.54 ± 0.07	0.70 ± 0.07
	With mass constraint							
μ	5280.08 ± 0.05	5280.04 ± 0.04	5280.14 ± 0.04	5280.06 ± 0.04	5280.09 ± 0.05	5280.04 ± 0.04	5280.14 ± 0.04	5280.06 ± 0.04
σ_{CB}	10.0 ± 0.7	12.2 ± 1.0	9.6 ± 0.6	8.9 ± 0.5	9.6 ± 0.7	12.1 ± 0.9	8.9 ± 0.7	7.9 ± 0.4
α_{CB1}	-1.91 ± 0.07	-1.79 ± 0.07	-1.89 ± 0.06	-1.92 ± 0.05	-1.78 ± 0.07	-1.63 ± 0.07	-1.74 ± 0.05	-1.78 ± 0.04
α_{CB2}	1.18 ± 0.21	0.77 ± 0.19	1.15 ± 0.16	1.24 ± 0.15	1.02 ± 0.20	0.70 ± 0.15	1.09 ± 0.15	1.18 ± 0.13
n_{CB1}	2.4	2.4	2.4	2.4	3.0	3.0	3.0	3.0
n_{CB2}	2.0	2.0	2.0	2.0	2.3	2.3	2.3	2.3
$\sigma_{Gauss}/\sigma_{CB}$	0.540 ± 0.018	0.493 ± 0.032	0.559 ± 0.018	0.586 ± 0.012	0.556 ± 0.021	0.498 ± 0.030	0.587 ± 0.019	0.595 ± 0.027
f_{CB1}	0.68 ± 0.10	0.72 ± 0.08	0.72 ± 0.07	0.71 ± 0.07	0.72 ± 0.09	0.72 ± 0.07	0.72 ± 0.06	0.73 ± 0.06
f_{DCB}	0.42 ± 0.07	0.23 ± 0.04	0.45 ± 0.08	0.55 ± 0.09	0.45 ± 0.08	0.228 ± 0.032	0.53 ± 0.10	0.71 ± 0.10

7.1.2 $\Lambda_b^0 \to pK^-\mu^+\mu^-$

As demonstrated in Sec. 5.2.3, there is a significant contribution from $\Lambda_b^0 \to J/\psi pK^-$ decays in the control mode dataset used in this analysis. It is reasonable to expect that the presence of the decay $\Lambda_b^0 \to J/\psi pK^-$ implies that the background decay $\Lambda_b^0 \to pK^-\mu^+\mu^-$ is present in the rare mode and that the invariant mass shapes of the two decays will be similar. The presence of the background in the signal region makes it unfeasible to apply a veto without significantly affecting the signal efficiency and so a component for this background is needed. The strategy for including this in the fit model is to extract the yield of the background and the shape of the component from the data. To select out these candidates in the dataset, PID requirements are applied to select out protons over pions. The optimised PID requirements on the other final state particles stay the same but a requirement of $\texttt{ProbNN}_p \times (1 - \texttt{ProbNN}_K) \times (1 - \texttt{ProbNN}_\pi) > 0.2$ is applied on the reconstructed pion.

To extract the yield of the $\Lambda_b^0 \to J/\psi pK^-$ in the control mode, a two dimensional fit is performed in $m(K^+\pi^-\mu^+\mu^-)$ and $m((\pi^- \to \bar{p})K^+\mu^+\mu^-)$, with a component for $B^0 \to J/\psi K^{*0}$ and $\Lambda_b^0 \to J/\psi pK^-$ in each reconstruction and sharing a common yield. The combinatorial background is relatively small and neglected in the fit. The component under the "correct" reconstruction for each decay is provided by the same double Crystal-Ball and Gaussian described above. The background component for each decay is given by a single Crystal-Ball, with all the parameters of the PDF allowed to float in the fit. The results of this fit, carried out separately for Run 1 and Run 2 data, are shown in Fig. 7.3.

Both the yields and line-shapes extracted by this method require the application of a PID cut to select out the candidates. For use in the full mass fit, these both need to be scaled to match the optimised PID requirements used to select the signal. The first step needed is to obtain a background subtracted sample of the $\Lambda_b^0 \to J/\psi pK^-$ mode. This is achieved by using the same *sPlot* method that is used to extract a $B^0 \to J/\psi K^{*0}$ sample in Sec. 5.1.3. A scaling of this background is then carried out making use of weights from PidCalib, with the ratio of weights between the two PID cuts applied on top of the extracted $sWeights$.

To fit the correct line-shape for the background, the ratio of weights needs to be applied on an event by event basis. A fit to these weighted candidates is then carried out, with the single component being the Crystal-Ball background component. The line-shape from this fit to the weighted background provides the component in the invariant mass fit, with the parametrisation given by:

$$B_{\Lambda_b^0}(m) = F_{CB}(m; \mu_{\Lambda_b^0}, \sigma^{CB}_{\Lambda_b^0}, \alpha_{\Lambda_b^0}, n_{\Lambda_b^0}). \tag{7.2}$$

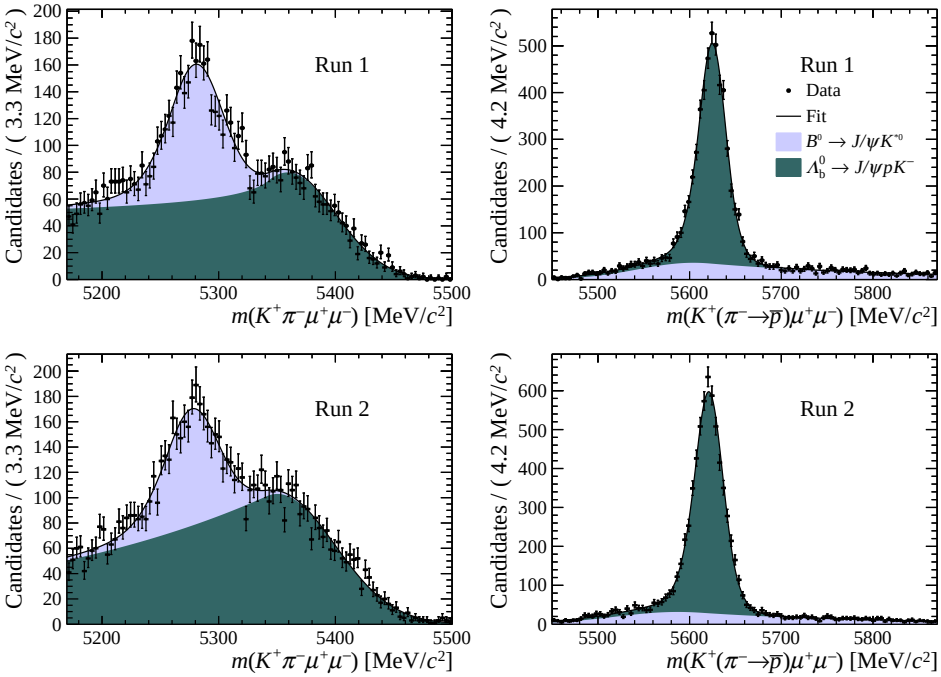

Figure 7.3: Fit to the $K^+\pi^-\mu^+\mu^-$ (left) and $K^+(\pi^- \to \bar{p})\mu^+\mu^-$ (right) invariant mass for $\Lambda_b^0 \to J/\psi p K^-$ candidates extracted from data. Fit components are detailed in the legend. Top row shows candidates from Run 1, bottom row candidates from Run 2.

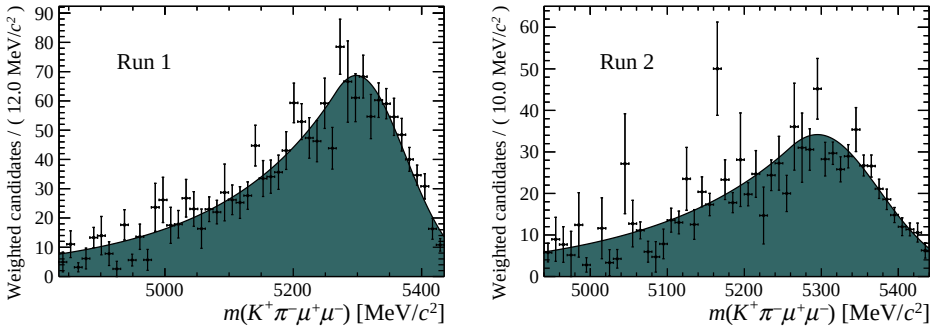

Figure 7.4: Fit to $K^+\pi^-\mu^+\mu^-$ invariant mass, *sWeighted* to select $\Lambda_b^0 \to J/\psi p K^-$ candidates with PID selection weights applied for Run 1 (left) and Run 2 (right).

The results of these fits for both Run 1 and Run 2 data are shown Fig. 7.4. The parameters of the fitted Crystal-Ball functions are given in Tab. 7.2. The fits to extract the $\Lambda_b^0 \to pK^-\mu^+\mu^-$ yield in the rare mode are shown in Fig. 7.5 and a summary of the fitted yields and the average ratio of PID weights for both the rare and control mode is given in Tab. 7.3.

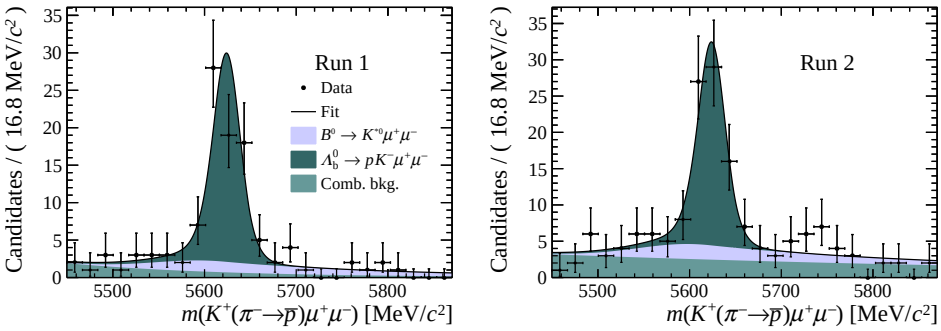

Figure 7.5: Fit to $K^+(\pi^- \to \bar{p})\mu^+\mu^-$ invariant mass for $\Lambda_b^0 \to pK^-\mu^+\mu^-$ candidates extracted from data in the rare mode q^2 region for Run 1 (left) and Run 2 (right). Fit components are detailed in the legend.

Table 7.2: Parameters of fit to $\Lambda_b^0 \to J/\psi pK^-$ candidates extracted from data with $sWeights$ applied and re-weighted for PID efficiency.

Parameter	Value	
	Run 1	Run 2
$\mu_{\Lambda_b^0}$	5298 ± 7	5294 ± 10
$\sigma_{\Lambda_b^0}$	76 ± 7	80 ± 9
$\alpha_{\Lambda_b^0}$	0.41 ± 0.05	0.45 ± 0.06
$n_{\Lambda_b^0}$	10 ± 4	10 ± 5

Table 7.3: Extracted $\Lambda_b^0 \to J/\psi pK^-$ and $\Lambda_b^0 \to pK^-\mu^+\mu^-$ yields from data as well as the relevant PID efficiencies to provide a constraint on the final number of these candidates in the control and rare mode fits.

		PID efficiency	Fitted yield	Fit constraint
Run 1	Control mode	0.175 ± 0.024	4705 ± 94	823 ± 114
	Rare mode		70 ± 8	12.3 ± 2.2
Run 2	Control mode	0.0351 ± 0.0019	5982 ± 95	210 ± 12
	Rare mode		69 ± 9	2.4 ± 0.3

7.1.3 $B^+ \to K^+\mu^+\mu^-$

The other major backgrounds that cannot be vetoed in this analysis are those from $B^+ \to J/\psi K^+$ and $B^+ \to K^+\mu^+\mu^-$ decays. This is because to application of a veto to remove these decays would distort the shape of the combinatorial background, as previously discussed in Sec. 5.2.4. A similar method of extracting a constraint

on the yield of the background and a line shape for the component as done for $\Lambda_b^0 \to pK^-\mu^+\mu^-$ decays can be used. In this case it is a simpler procedure as there is no particle mis-identification involved with this background and thus no need to apply PID requirements to select it out. In Sec. 5.2.4 it has been demonstrated that this decay can be clearly identified by analysing the $m(K^+\mu^+\mu^-)$ invariant mass. The yield of the decay mode can be extracted simply by fitting this invariant mass and the line shape can be extracted by fitting the $m(K^+\pi^-\mu^+\mu^-)$ mass of candidates in the region $5220 < m(K^+\mu^+\mu^-) < 5350$ MeV/c^2. Fits to both of these are carried out in bins of neural network response as the background efficiency is heavily correlated with the response.

Fits to all four bins of neural network response for both Run 1 and Run 2 data are carried out, with a component for the $B^+ \to J/\psi K^+$ candidates included using a double Crystal-Ball and Gaussian PDF and an exponential function to account for combinatorial background. In the $m(K^+\mu^+\mu^-)$ region these fits are carried out over, no contribution from the signal decays $B^0 \to J/\psi K^{*0}$ or $B_s^0 \to J/\psi \overline{K}^{*0}$ is expected. These fits are shown in Fig. 7.6 with the resulting yields and thus the constraints applied on the background yields in the full mass fit shown in Tab. 7.4.

Table 7.4: Yields of $B^+ \to J/\psi K^+$ candidates extracted from data in bins of neural network response for Run 1 and Run 2 data.

		Bin 1	Bin 2	Bin 3	Bin 4
$N(B^+ \to J/\psi K^+)$	Run 1	3749 ± 62	912 ± 31	231 ± 18	18 ± 5
	Run 2	5140 ± 77	984 ± 35	339 ± 20	51 ± 7

The PDFs used to describe the $B^+ \to J/\psi K^+$ background in each bin of neural network response are Crystal-Ball functions, the full parameterisations of which are given by:

$$B_{K\mu\mu}(m) = F_{CB}(m; \mu_{K\mu\mu}, \sigma_{K\mu\mu}, \alpha_{K\mu\mu}, n_{K\mu\mu}). \tag{7.3}$$

The low yield of the 4th bin in both the Run 1 and Run 2 datasets make it impossible to fit and extract the parameters for the background component. As such, the final two bins are merged and the line-shape from these combined bins is used for both the 3rd and 4th bin. It is also necessary to extract these components separately for the fit to the control mode with and without a constraint on the J/ψ mass as the constraint significantly affects the shape of the background. Figure 7.7 shows these fits without a J/ψ mass constraint, Fig. 7.8 with a J/ψ mass constraint and the

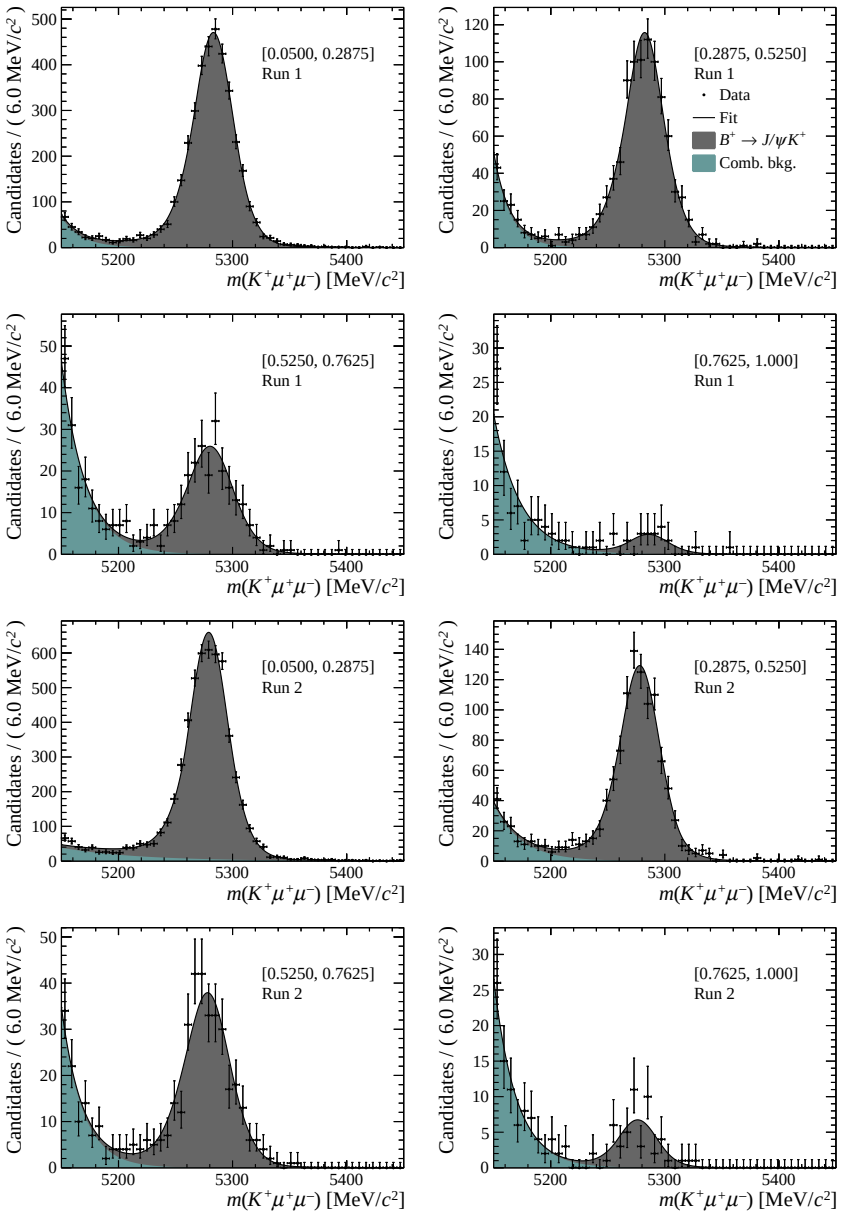

Figure 7.6: Fits to the $K^+\mu^+\mu^-$ invariant mass of $B^+ \to J/\psi K^+$ candidates extracted from data to provide constraints on candidate yields for the Run 1 (top four) and Run 2 (bottom four) data sets. The candidates are divided into four independent bins of increasing neural network response per data taking period. Fit components are detailed in the legend.

resulting parameters from these fits are given in Tab. 7.5.

The background from $B^+ \to J/\psi K^+$ is clearly a significant component of the fit in the control mode and is vital for the best quality fit. However, when the same procedure is carried out to extract a yield for the rare mode there is no visible

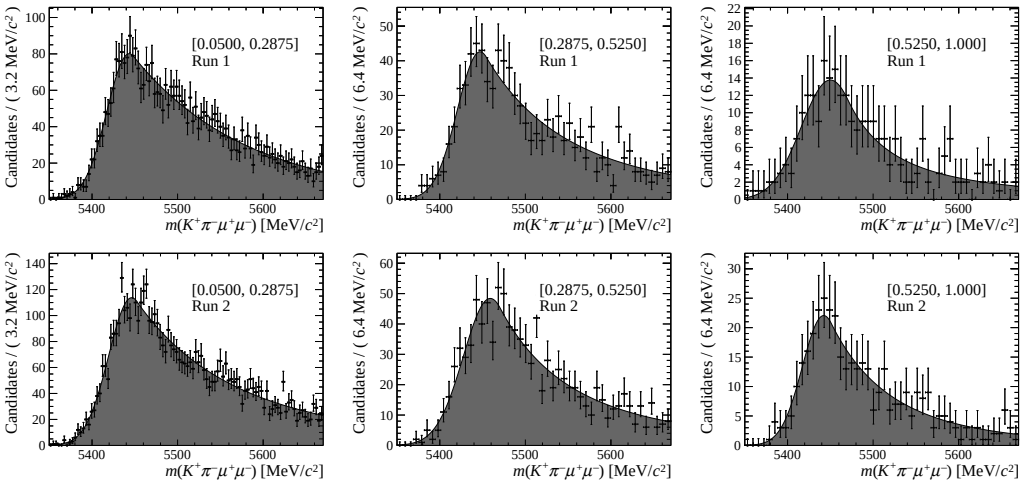

Figure 7.7: Fits to $K^+\pi^-\mu^+\mu^-$ invariant mass of $B^+ \to J/\psi K^+$ candidates extracted from data for the Run 1 (top three) and Run 2 (bottom three) data sets. The candidates are divided into independent bins of neural network response where the first bin (left), second bin (middle) and final two bins combined (right) are shown in increasing order of output purity.

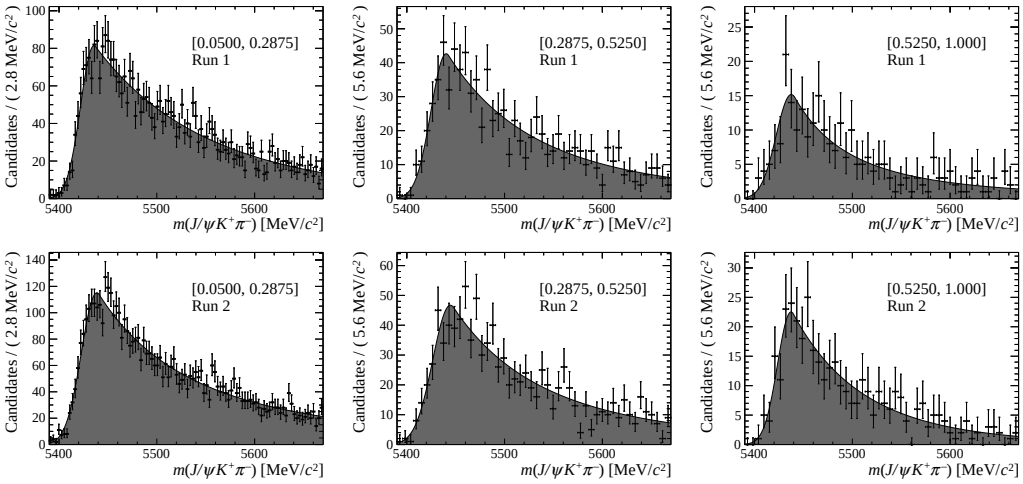

Figure 7.8: Fits to $J/\psi K^+\pi^-$ invariant mass of $B^+ \to J/\psi K^+$ candidates extracted from data for the Run 1 (top three) and Run 2 (bottom three) data sets, where a constraint on the J/ψ mass has been applied. The candidates are divided into independent bins of neural network response where the first bin (left), second bin (middle) and final two bins combined (right) are shown in increasing order of output purity.

contribution from $B^+ \to K^+\mu^+\mu^-$ decays and no component for it is included in the fit to the rare mode.

Table 7.5: Parameters of fits to $K^+\pi^-\mu^+\mu^-$ invariant mass for $B^+ \to J/\psi K^+$ candidates extracted from data in bins of neural network response for both Run 1 and Run 2 data.

Parameter	Run 1			Run 2		
	Bin 1	Bin 2	Bin 3+4	Bin 1	Bin 2	Bin 3+4
	No mass constraint					
μ	5443.5 ± 1.9	5447 ± 4	5450 ± 6	5446.3 ± 1.9	5459 ± 4	5442 ± 4
σ	26.1 ± 1.2	26.6 ± 2.4	34 ± 4	26.9 ± 1.1	31.5 ± 2.6	24.9 ± 2.6
α	-0.194 ± 0.013	-0.27 ± 0.09	-0.61 ± 0.28	-0.229 ± 0.035	-0.37 ± 0.12	-0.35 ± 0.12
n	100 ± 51	4 ± 5	1.9 ± 1.9	5 ± 4	3.2 ± 3.3	5 ± 7
	With mass constraint					
μ	5438.2 ± 1.6	5445 ± 5	5438 ± 4	5438.2 ± 1.6	5445 ± 5	5437 ± 4
σ	15.6 ± 0.9	18.5 ± 2.9	15.4 ± 2.8	15.6 ± 0.9	18.5 ± 2.9	13.2 ± 2.4
α	-0.155 ± 0.025	-0.19 ± 0.07	-0.27 ± 0.11	-0.155 ± 0.025	-0.19 ± 0.07	-0.18 ± 0.07
n	2.8 ± 1.2	4 ± 5	2.4 ± 1.8	2.8 ± 1.2	4 ± 5	13 ± 53

7.1.4 $B^0 \to K^{*0}\mu^+\mu^-$ as a background to $B^0 \to J/\psi K^{*0}$

In the fit to the $B^0 \to J/\psi K^{*0}$ control mode there will be some contribution from $B^0 \to K^{*0}\mu^+\mu^-$ decays that lie in the q^2 region used to select decays via the J/ψ resonance. Where no constraint on the J/ψ mass has been applied, this background is indistinguishable from the $B^0 \to J/\psi K^{*0}$ signal. However, when a mass constraint is applied this is no longer true. As the $B^0 \to K^{*0}\mu^+\mu^-$ does not decay via a J/ψ resonance, the mass constraint does not improve the resolution mode but instead acts to smear the reconstructed mass of the selected candidates, with the difference between the J/ψ mass and dimuon mass of the $B^0 \to K^{*0}\mu^+\mu^-$ decay being the amount the candidate is shifted by. This causes a broad shape which is distinct from the combinatorial background.

A component for $B^0 \to K^{*0}\mu^+\mu^-$ decays in the control mode is added by taking the shape from simulated $B^0 \to K^{*0}\mu^+\mu^-$ candidates with a mass constraint applied and in the relevant q^2 region. This is modelled using a non-parametric kernel density estimate, implemented as a RooKeysPDF, leaving the only free parameter as the yield of the component:

$$B_{K^{*0}\mu^+\mu^-}(m) = F_{Keys}(m). \tag{7.4}$$

This yield is constrained in the control mode by scaling the yield of the $B^0 \to K^{*0}\mu^+\mu^-$ candidates in the rare mode to the q^2 of the control mode.

The candidates from simulation and the resulting PDFs used to model them for both Run 1 and Run 2 are shown in Fig. 7.9. Note that unlike other components in the fit taken from simulation, no attempt is made to account for any differences between data and simulation and the shape is used directly. There may be some small differences, however as the yield of these decays is small with respect to the

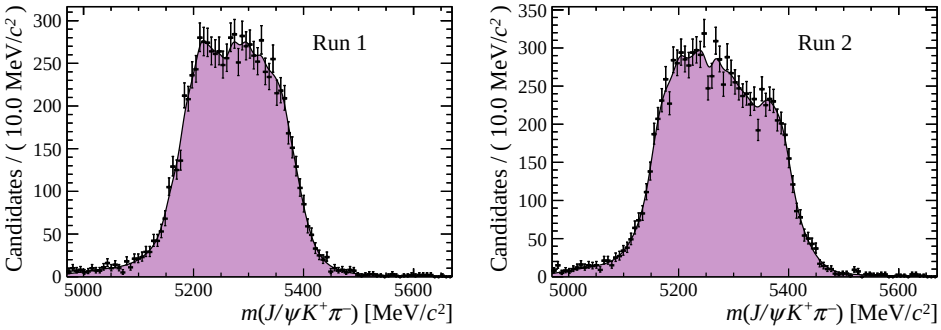

Figure 7.9: Fit to $J/\psi K^+\pi^-$ invariant mass of $B^0 \to K^{*0}\mu^+\mu^-$ candidates from simulation in the q^2 window of the control mode for Run 1 (left) and Run 2 (right).

$B^0 \to J/\psi K^{*0}$ decay any effect on the final result will be small.

7.1.5 Kaon-pion misidentification

In Sec. 5.2.3 it was shown that a background from $B^0 \to K^{*0}\mu^+\mu^-$ decays where the final state kaon has been misidentified as a pion and vice versa is present in the dataset. This background is expected to be heavily suppressed by the application of PID requirements, however a small component can still exist which acts as a modification of the $B^0 \to K^{*0}\mu^+\mu^-$ line shape. As this component is a background with respect to the $B^0_s \to \overline{K}^{*0}\mu^+\mu^-$, even though the yield is small it is important for it to be correctly modelled. Note that the corresponding background from $B^0_s \to \overline{K}^{*0}\mu^+\mu^-$ with the swapped kaon and pion identification is neglected as it is expected to be approximately 100 times smaller. The line shape of these candidates can be taken from $B^0 \to J/\psi K^{*0}$ simulation where the true pion is associated with the reconstructed kaon and vice-versa. The sum of two Crystal-Ball functions is used to model this background, with the two components sharing a mean but with different widths. This line-shape is parameterised as:

$$\begin{aligned} B_{K\pi\text{swap}}(m) =& f \cdot F_{CB}(m; \mu_{K\pi\text{swap}}, \sigma_{CB1}, \alpha_{CB1}, n_{CB1}) + \\ & (1-f) \cdot F_{CB}(m; \mu_{K\pi\text{swap}}, \sigma_{CB2}, \alpha_{CB2}, n_{CB2}) \end{aligned} \tag{7.5}$$

A separate component for this background is used in each bin of neural network response, with the width of the first Crystal-Ball function fixed relative to

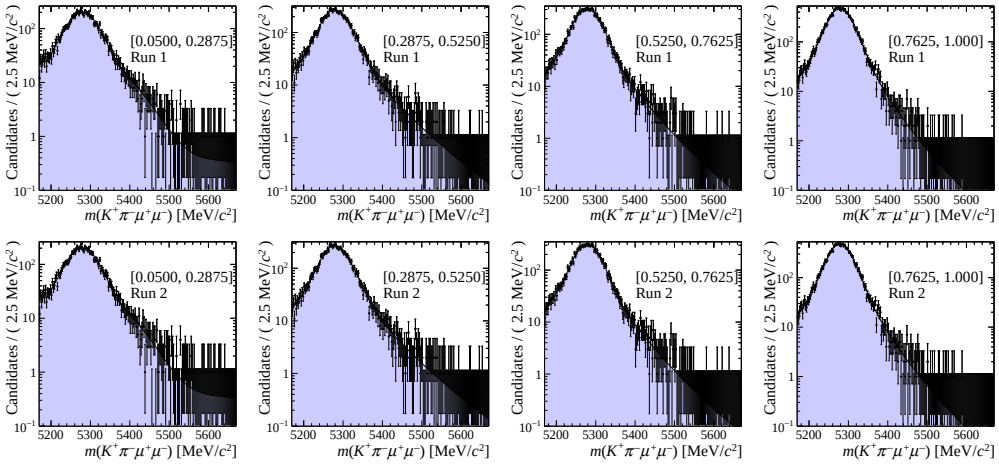

Figure 7.10: Fit in bins of increasing neural network response to $B^0 \to J/\psi K^{*0}$ background where the kaon has been misidentified as a pion and vice-versa. The background is taken from simulation and is shown for Run 1 (top four) and Run 2 (bottom four). Results are shown on a log scale.

the core width of the signal component as determined in Sec. 7.1.1. The yield of this component is also directly constrained to the yield of $B^0 \to J/\psi K^{*0}/B^0 \to K^{*0}\mu^+\mu^-$ candidates by making use of the ratio of PID efficiencies determined from PIDCALIB for both components as taken from simulation. There are no free parameters for this component in the full mass fit, all of the PDF parameters are based on the B^0 component. The fits to this background component taken from simulation are shown in Fig. 7.10 with the parameters of this fit shown in Tab. 7.6.

Table 7.6: Parameters of background PDFs obtained from fits to $B^0 \to J/\psi K^{*0}$ kaon-pion swap background from simulation in bins of neural network response for both Run 1 and Run 2. Labels 1 and 2 distinguish between the two CB components.

Parameter	Bin 1	Bin 2	Bin 3	Bin4
	Run 1			
μ	5278.3 ± 1.1	5278.6 ± 0.6	5279.1 ± 0.8	5278.5 ± 0.6
σ_{CB1}	31.7 ± 2.7	26 ± 4	31.1 ± 1.7	26.2 ± 1.5
$\sigma_{CB2}/\sigma_{CB1}$	1.4 ± 0.5	2.15 ± 0.18	0.96 ± 0.10	1.18 ± 0.10
α_{CB1}	-1.3 ± 0.4	-2.95 ± 0.19	-0.94 ± 0.10	-1.00 ± 0.14
α_{CB2}	0.6 ± 0.5	1.0 ± 0.4	0.66 ± 0.13	0.83 ± 0.15
n_{CB1}	2.9 ± 2.9	0.1 ± 0.4	10 ± 7	10 ± 6
n_{CB2}	10 ± 5	10 ± 9	10 ± 9	10 ± 9
f_{CB1}	0.68 ± 0.18	0.53 ± 0.17	0.53 ± 0.10	0.51 ± 0.12
	Run 2			
μ	5279.2 ± 0.6	5276.1 ± 1.1	5278.0 ± 1.0	5278.48 ± 0.30
σ_{CB1}	30.8 ± 1.3	18 ± 4	30 ± 7	25.7 ± 0.9
$\sigma_{CB2}/\sigma_{CB1}$	2.03 ± 0.10	2.0 ± 0.5	1.04 ± 0.31	1.89 ± 0.09
α_{CB1}	-2.75 ± 0.11	-0.36 ± 0.09	-0.70 ± 0.26	-2.7 ± 0.4
α_{CB2}	0.84 ± 0.18	1.14 ± 0.11	1.19 ± 0.25	1.16 ± 0.16
n_{CB1}	0.10 ± 0.18	10 ± 8	10 ± 10	0.6 ± 0.5
n_{CB2}	10 ± 10	10 ± 9	2.6 ± 2.8	10 ± 9
f_{CB1}	0.55 ± 0.06	0.25 ± 0.04	0.31 ± 0.10	0.62 ± 0.05

7.1.6 Combinatorial background

Combinatorial background is modelled in the full fit by independent exponential components in each of the 4 neural network bins for the two datasets. Both the slope parameter and yield of each exponential component is allowed to float in the fit. This is done separately for the fit to the control mode and to the rare mode. In the case where the yield of the combinatorial is consistent with zero in

an individual neural network bin, the slope parameter is fixed to be zero, *i.e.* it is assumed the background is constant over the invariant mass distribution. This is due to the fit becoming unstable when the combinatorial yield is low, requiring the slope parameter to be fixed for the fit the converge.

7.2 Full fit

The full fit model is constructed from the individual components from the previous section, with some parameters of the individual PDFs fixed and some allowed to float. In addition some terms are constrained using external information allowing the parameter to float within an uncertainty. The fit is carried out in eight bins of neural network response, four each for Run 1 and Run 2 data. In each bin of NN response i, the full PDF is give by:

$$\begin{aligned}\mathrm{PDF}^{(i)}(m_{K\pi\mu\mu}) =& f^{(i)}_{B^0} \cdot S^{(i)}_{B^0}(m_{K\pi\mu\mu}) \; + \; f^{(i)}_{B^0_s} \cdot S^{(i)}_{B^0_s}(m_{K\pi\mu\mu}) + \\ & f^{(i)}_{\Lambda^0_b} \cdot B^{(i)}_{\Lambda^0_b}(m_{K\pi\mu\mu}) \; + \; f^{(i)}_{K\pi\,\mathrm{swap}} \cdot B^{(i)}_{K\pi\,\mathrm{swap}}(m_{K\pi\mu\mu}) + \\ & f^{(i)}_{K\mu\mu} \cdot B^{(i)}_{K\mu\mu}(m_{K\pi\mu\mu}) \; + \; f^{(i)}_{\mathrm{comb}} \cdot B^{(i)}_{\mathrm{comb}}(m_{K\pi\mu\mu})\end{aligned} \tag{7.6}$$

where the fraction of component k in bin i is

$$f^{(i)}_k = \frac{N^{(i)}_k}{\sum\limits_k N^{(i)}_k} \tag{7.7}$$

For the J/ψ mass constrained fit of the $B^0 \to J/\psi K^{*0}$ control mode an extra component for $B^0 \to K^{*0}\mu^+\mu^-$ background decays is included:

$$f^{(i)}_{K^{*0}\mu\mu} \cdot B^{(i)}_{K^{*0}\mu\mu}(m_{K\pi\mu\mu}) \tag{7.8}$$

Expanding the PDF to show the fit parameters for each bin

$$\begin{aligned}\mathrm{PDF}^{(i)}(m) =& f^{(i)}_{B^0} \cdot S^{(i)}_{B^0}(m; \mu_{B^0}, \sigma^{CB(i)}_{B^0}, \sigma^{Gauss(i)*}_{B^0}, \alpha^{\mathrm{scale\,1}}_{B^0}, \alpha^{\mathrm{scale\,2}}_{B^0}, \\ & \alpha^{CB1(i)*}_{B^0}, n^{CB1(i)*}_{B^0}, \alpha^{CB2(i)*}_{B^0}, n^{CB2(i)*}_{B^0}, f^{CB1(i)*}_{B^0}, f^{DCB(i)*}_{B^0}) + \\ & f^{(i)}_{B^0_s} \cdot S^{(i)}_{B^0_s}(m; \mu^*_{B^0_s}, \sigma^{CB(i)}_{B^0}, \sigma^{Gauss(i)*}_{B^0}, \alpha^{\mathrm{scale\,1}}_{B^0}, \alpha^{\mathrm{scale\,2}}_{B^0}, \\ & \alpha^{CB1(i)*}_{B^0}, n^{CB1(i)*}_{B^0}, \alpha^{CB2(i)*}_{B^0}, n^{CB2(i)*}_{B^0}, f^{CB1(i)*}_{B^0}, f^{DCB(i)*}_{B^0}) + \\ & f^{(i)}_{\Lambda^0_b} \cdot B_{\Lambda^0_b}(m; \mu^*_{\Lambda^0_b}, \sigma^*_{\Lambda^0_b}, \alpha^*_{\Lambda^0_b}, n^*_{\Lambda^0_b}) + \\ & f^{(i)}_{K\mu\mu} \cdot B_{K\mu\mu}(m; \mu^{(i)*}_{K\mu\mu}, \sigma^{(i)*}_{K\mu\mu}, \alpha^{(i)*}_{K\mu\mu}, n^{(i)*}_{K\mu\mu}) + \\ & f^{(i)}_{K\pi\,\mathrm{swap}} \cdot B^{(i)}_{K\pi\,\mathrm{swap}}(m; \mu_{B^0}, \sigma^{CB1(i)*}_{K\pi\,\mathrm{swap}}, \sigma^{CB2(i)*}_{K\pi\,\mathrm{swap}}, \\ & \alpha^{CB1(i)*}_{K\pi\,\mathrm{swap}}, n^{CB1(i)*}_{K\pi\,\mathrm{swap}}, \alpha^{CB2(i)*}_{K\pi\,\mathrm{swap}}, n^{CB2(i)*}_{K\pi\,\mathrm{swap}}, f^{CB1(i)*}_{K\pi\,\mathrm{swap}}) + \\ & f^{(i)}_{\mathrm{comb}} \cdot B^{(i)}_{\mathrm{comb}}(m; \alpha^{(i)})\end{aligned} \tag{7.9}$$

The fractions are expressed as

$$
\begin{aligned}
f^{(i)}_{B^0} &= \frac{\kappa^{(j)}_{B^0}\epsilon^{(i)}N_{B^0}}{N^{(i)}},\ f^{(i)}_{B^0_s} = \frac{\kappa^{(j)}_{B^0_s}\epsilon^{(i)}N_{B^0_s}}{N^{(i)}} \\
f^{(i)}_{\Lambda^0_b} &= \frac{\epsilon^{(i)}N^{(j)\dagger}_{\Lambda^0_b}}{N^{(i)}},\ f^{(i)}_{K\pi\,\mathrm{swap}} = Y^{*}_{K\pi\,\mathrm{swap}}\frac{\kappa^{(j)}_{B^0}\epsilon^{(i)}N_{B^0}}{N^{(i)}} \\
f^{(i)}_{K\mu\mu} &= \frac{N^{(i)\dagger}_{K\mu\mu}}{N^{(i)}},\ f^{(i)}_{\mathrm{comb}} = \frac{N^{(i)}_{\mathrm{comb}}}{N^{(i)}}
\end{aligned}
\tag{7.10}
$$

where $\kappa^{(j)}$ is the fraction of B^0/B^0_s candidates in either the Run 1 or Run 2 dataset, $\epsilon^{(i)}$ is the fraction of the signal candidates in each dataset in bin i and $Y_{K\pi\,\mathrm{swap}}$ is the scaling of the kaon-pion swapped background yields derived from PidCalib. Variables labeled $(*)$ are fixed (mostly to results from MC) and variables labeled $(\dagger)$ indicate that the corresponding yield is Gaussian constrained to results from yield extraction fits. The yield of Λ^0_b candidates is constrained separately for Run 1 and Run 2 datasets. The width of $\sigma^{Gauss}_{B^0}$ is fixed relative to $\sigma^{CB}_{B^0}$ from MC. The width of $\sigma^{CB\{1,2\}}_{K\pi\,\mathrm{swap}}$ is also fixed relative to $\sigma^{CB}_{B^0}$, *i.e.* so worse resolution in data compared to the result from MC would also smear this distribution. The parameter $\alpha^{\mathrm{scale}(1/2)}$ is a parameter common across all bins which multiplies the corresponding $\alpha^{CB(1/2)}_{B^0/B^0_s}$ parameters in individual neural network bins and is allowed to float in the control mode.

The full fit PDF for the rare mode is given by

$$
\begin{aligned}
\mathrm{PDF}^{(i)}(m) =& f^{(i)}_{B^0}\cdot S^{(i)}_{B^0}(m;\mu^{\ddagger}_{B^0},\sigma^{CB(i)\ddagger}_{B^0},\sigma^{Gauss(i)\ddagger}_{B^0},\alpha^{\mathrm{scale\,1}\ddagger}_{B^0},\alpha^{\mathrm{scale\,2}\ddagger}_{B^0}, \\
&\qquad \alpha^{CB1(i)*}_{B^0},n^{CB1(i)*}_{B^0},\alpha^{CB2(i)*}_{B^0},n^{CB2(i)*}_{B^0},f^{CB1(i)*}_{B^0},f^{DCB(i)*}_{B^0}) + \\
& f^{(i)}_{B^0_s}\cdot S^{(i)}_{B^0_s}(m;\mu^{*}_{B^0_s},\sigma^{CB(i)\ddagger}_{B^0},\sigma^{Gauss(i)*}_{B^0},\alpha^{\mathrm{scale\,1}\ddagger}_{B^0},\alpha^{\mathrm{scale\,2}\ddagger}_{B^0}, \\
&\qquad \alpha^{CB1(i)*}_{B^0},n^{CB1(i)*}_{B^0},\alpha^{CB2(i)*}_{B^0},n^{CB2(i)*}_{B^0},f^{CB1(i)*}_{B^0},f^{DCB(i)*}_{B^0}) + \\
& f^{(i)}_{\Lambda^0_b}\cdot B_{\Lambda^0_b}(m;\mu^{*}_{\Lambda^0_b},\sigma^{*}_{\Lambda^0_b},\alpha^{*}_{\Lambda^0_b},n^{*}_{\Lambda^0_b}) + \\
& f^{(i)}_{K\pi\mathrm{swap}}\cdot B^{(i)}_{K\pi\,\mathrm{swap}}(m;\mu_{B^0},\sigma^{CB1(i)*}_{K\pi\,\mathrm{swap}},\sigma^{CB2(i)*}_{K\pi\,\mathrm{swap}}, \\
&\qquad \alpha^{CB1(i)*}_{K\pi\,\mathrm{swap}},n^{CB1(i)*}_{K\pi\,\mathrm{swap}},\alpha^{CB2(i)*}_{K\pi\,\mathrm{swap}},n^{CB2(i)*}_{K\pi\,\mathrm{swap}},f^{CB1(i)*}_{K\pi\,\mathrm{swap}}) + \\
& f^{(i)}_{\mathrm{comb}}\cdot B^{(i)}_{\mathrm{comb}}(m;\alpha^{(i)})
\end{aligned}
\tag{7.11}
$$

Variables labelled $(\ddagger)$ indicate variables that have been fixed to their values from the fit to the control mode. The key differences between the fits to the rare mode and the control mode are that no component for $B^+\to K^+\mu^+\mu^-$ decays is included

in the rare mode, and that in the rare mode the fraction $\epsilon^{(i)}$ is fixed to its value from MC (which is 1/4 by construction of the binning scheme).

7.3 Fit to data

Before fitting the rare mode for the $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ yield, it is necessary to perform a fit to the $B^0 \to J/\psi K^{*0}$ control mode without a constraint on the J/ψ to extract the scaling for the signal resolution and tail parameters to account for data-simulation differences. This fit, shown in individual bins of neural network response, is shown in Fig. 7.11 with the resulting fit parameters shown in Tab. 7.7. The slope parameter of the combinatorial component is fixed to zero in the final two bins for the Run 1 dataset where the combinatorial background yield is consistent with zero, as discussed in Sec. 7.1.6. The fit to the control mode with a J/ψ mass constraint is also carried out. As detailed above, this has the same components as the fit without a mass constraint but with the addition of a component for $B^0 \to K^{*0}\mu^+\mu^-$ background decays. Figure 7.12 shows the data and results of the fit and the fit parameters are given in Tab. 7.8

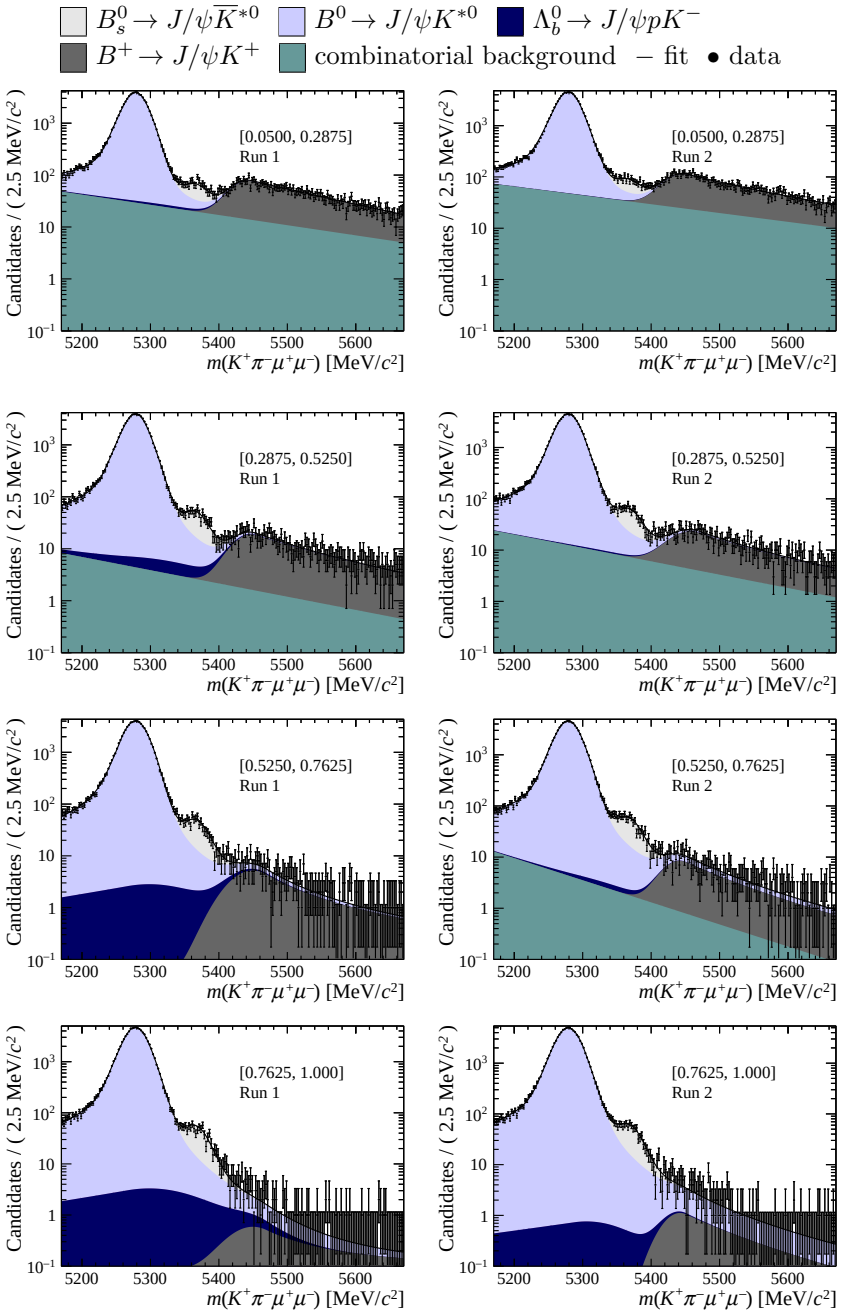

Figure 7.11: Distribution of reconstructed $K^+\pi^-\mu^+\mu^-$ invariant mass of candidates in the control mode without a constraint on the J/ψ mass for the Run 1 (left) and Run 2 (right) data sets as well as the result of a maximum likelihood fit. The candidates are divided into four independent bins of increasing neural network response per data taking period.

Table 7.7: Parameters of invariant mass fit to the $B^0 \to J/\psi K^{*0}$ control mode in bins of neural network response for both the Run 1 and Run 2 datasets. Superscript number labels distinguish between parameters for each neural network bin.

Parameter	Value	
	Run 1	Run 2
μ_{B^0}	5278.00 ± 0.04	5278.66 ± 0.04
α_1^{scale}	0.946 ± 0.023	0.876 ± 0.012
α_2^{scale}	0.989 ± 0.008	0.994 ± 0.010
$\sigma_{B^0}^1$	13.95 ± 0.06	14.51 ± 0.06
$\sigma_{B^0}^2$	13.29 ± 0.06	15.18 ± 0.07
$\sigma_{B^0}^3$	15.69 ± 0.07	14.62 ± 0.06
$\sigma_{B^0}^4$	14.62 ± 0.07	15.56 ± 0.07
Exp. slope (Bin 1)	-0.00451 ± 0.00029	-0.00396 ± 0.00021
Exp. slope (Bin 2)	-0.0059 ± 0.0015	-0.0060 ± 0.0007
Exp. slope (Bin 3)	0.0	-0.0086 ± 0.0013
Exp. slope (Bin 4)	0.0	-0.015 ± 0.010
f_{sig}^1	0.2349 ± 0.0008	0.2450 ± 0.0008
f_{sig}^2	0.2360 ± 0.0009	0.2412 ± 0.0008
f_{sig}^3	0.2451 ± 0.0008	0.2448 ± 0.0008
f_{sig}^4	0.2840 ± 0.0008	0.2690 ± 0.0008
$N(B^+ \to J/\psi K^+)$ (Bin 1)	3675 ± 56	5100 ± 68
$N(B^+ \to J/\psi K^+)$ (Bin 2)	920 ± 27	1000 ± 31
$N(B^+ \to J/\psi K^+)$ (Bin 3)	251 ± 10	338 ± 19
$N(B^+ \to J/\psi K^+)$ (Bin 4)	24.3 ± 2.1	50 ± 4
N(Comb. bkg.) (Bin 1)	3814 ± 172	6369 ± 222
N(Comb. bkg.) (Bin 2)	534 ± 122	1520 ± 158
N(Comb. bkg.) (Bin 3)	24 ± 32	559 ± 124
N(Comb. bkg.) (Bin 4)	19 ± 14	5 ± 26
$N(\Lambda_b^0 \to J/\psi p K^-)$	888 ± 69	210 ± 12
$\kappa_{B^0}^1$	0.4688 ± 0.0007	
$\kappa_{B_s^0}^1$	0.460 ± 0.013	
$N(B^0 \to J/\psi K^{*0})$	628804 ± 916	
$N(B_s^0 \to J/\psi \overline{K}^{*0})$	5854 ± 154	

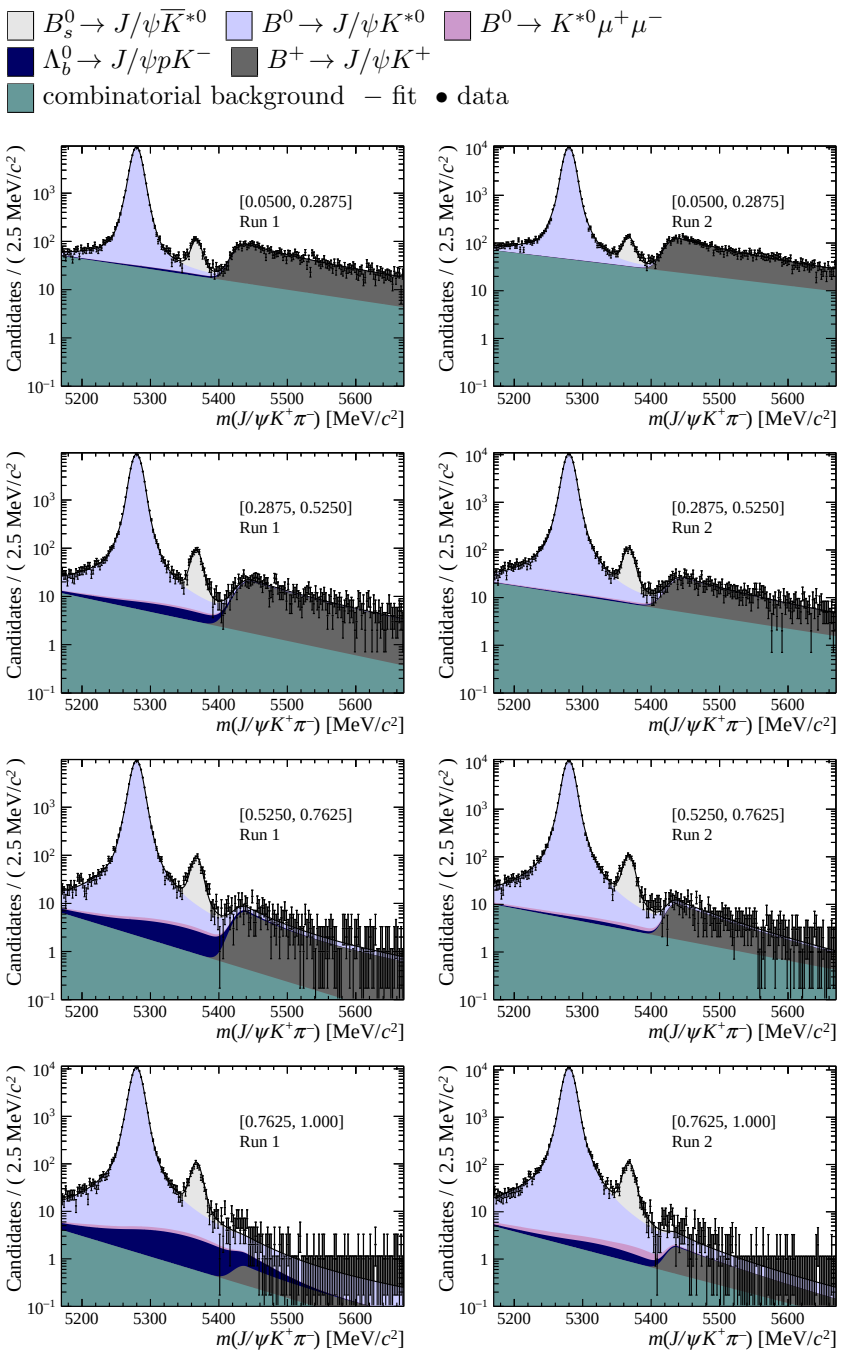

Figure 7.12: Distribution of reconstructed $K^+\pi^-\mu^+\mu^-$ invariant mass of candidates in the control mode with a constraint on the J/ψ mass for the Run 1 (left) and Run 2 (right) data sets as well as the result of a maximum likelihood fit. The candidates are divided into four independent bins of increasing neural network response per data taking period.

Table 7.8: Parameters of invariant mass fit to the $B^0 \to J/\psi K^{*0}$ control mode with an mass constraint on the J/ψ applied in bins of neural network response for both the Run 1 and Run 2 datasets. Superscript number labels distinguish between parameters for each neural network bin.

Parameter	Value	
	Run 1	Run 2
μ_{B^0}	5279.605 ± 0.016	5279.794 ± 0.015
α_1^{scale}	0.947 ± 0.007	0.926 ± 0.008
α_2^{scale}	0.988 ± 0.018	0.971 ± 0.018
$\sigma_{B^0}^1$	11.15 ± 0.04	14.03 ± 0.05
$\sigma_{B^0}^2$	13.75 ± 0.05	10.38 ± 0.04
$\sigma_{B^0}^3$	10.79 ± 0.04	11.61 ± 0.04
$\sigma_{B^0}^4$	9.901 ± 0.034	12.85 ± 0.04
Exp. slope (Bin 1)	-0.00490 ± 0.00027	-0.00396 ± 0.00018
Exp. slope (Bin 2)	-0.0069 ± 0.0010	-0.0052 ± 0.0005
Exp. slope (Bin 3)	-0.011 ± 0.003	-0.0063 ± 0.0008
Exp. slope (Bin 4)	-0.047 ± 0.030	-0.0089 ± 0.0016
f_{sig}^1	0.2346 ± 0.0008	0.2447 ± 0.0008
f_{sig}^2	0.2361 ± 0.0008	0.2414 ± 0.0008
f_{sig}^3	0.2450 ± 0.0008	0.2451 ± 0.0008
f_{sig}^4	0.2843 ± 0.0008	0.2688 ± 0.0008
$N(B^+ \to J/\psi K^+)$ (Bin 1)	3664 ± 53	5046 ± 66
$N(B^+ \to J/\psi K^+)$ (Bin 2)	904 ± 26	980 ± 30
$N(B^+ \to J/\psi K^+)$ (Bin 3)	235 ± 11	318 ± 16
$N(B^+ \to J/\psi K^+)$ (Bin 4)	16.2 ± 2.0	49 ± 5
N(Comb. bkg.) (Bin 1)	3718 ± 129	6099 ± 152
N(Comb. bkg.) (Bin 2)	640 ± 80	1472 ± 106
N(Comb. bkg.) (Bin 3)	226 ± 68	603 ± 84
N(Comb. bkg.) (Bin 4)	67 ± 68	226 ± 65
$N(\Lambda_b^0 \to J/\psi p K^-)$	830 ± 45	221 ± 12
$N(B^0 \to K^{*0}\mu^+\mu^-)$	212 ± 6	255 ± 7
$\kappa_{B^0}^1$	0.4680 ± 0.0007	
$\kappa_{B_s^0}^1$	0.464 ± 0.008	
$N(B^0 \to J/\psi K^{*0})$	627244 ± 837	
$N(B_s^0 \to J/\psi \overline{K}^{*0})$	5730 ± 94	

The final fit to be carried out to the data is to the rare $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ mode. In this fit the number of free parameters is significantly smaller, with the mean, widths and tail parameter scalings of the signal components fixed to the first control mode fit and the signal efficiency in each bin being 1/4 by construction. The relative fraction of B^0 and B_s^0 in the Run 1 and Run 2 datasets are also fixed to their values from the control mode fit. The results of the fit are shown in Fig. 7.13 with the fit parameters shown in Tab. 7.9. Once again the slope parameter of one of the combinatorial components, this time the 4th bin of the Run 2 dataset, has been fixed to zero for purposes of fit convergence when the background yield is consistent with zero.

Table 7.9: Parameters of invariant mass fit to the $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ candidates in bins of neural network response for both the Run 1 and Run 2 datasets.

Parameter	Value	
	Run 1	Run 2
Exp. slope (Bin 1)	-0.00246 ± 0.00029	-0.00216 ± 0.00021
Exp. slope (Bin 2)	-0.0012 ± 0.0011	-0.0021 ± 0.0007
Exp. slope (Bin 3)	-0.001 ± 0.005	-0.0043 ± 0.0023
Exp. slope (Bin 4)	-0.011 ± 0.003	0.0
N(Comb. bkg.) (Bin 1)	717 ± 31	1350 ± 40
N(Comb. bkg.) (Bin 2)	76 ± 13	152 ± 17
N(Comb. bkg.) (Bin 3)	8 ± 6	27 ± 12
N(Comb. bkg.) (Bin 4)	23 ± 9	3 ± 3
$N(\Lambda_b^0 \to pK^-\mu^+\mu^-)$	12.3 ± 2.2	2.44 ± 0.34
$N(B^0 \to K^{*0}\mu^+\mu^-)$	4157 ± 72	
$N(B_s^0 \to \overline{K}^{*0}\mu^+\mu^-)$	38 ± 12	

The fit to the rare mode results in a yield of 38 ± 12 signal decay, summing over all 8 independent datasets. The contribution in any individual bin is too small to be clearly seen in the projections of the fit and so a combination of the bins is produced to highlight the contribution. Fig. 7.14 shows this combination, where the three most signal-like bins of neural network response in each dataset have been merged. This is not a separate fit to a combined dataset but instead the addition of the individual datasets overlaid with the addition of the resulting fits. The least signal-like bin in each dataset has been excluded in this combination as this bin has a significantly higher combinatorial background that would act to obscure the signal.

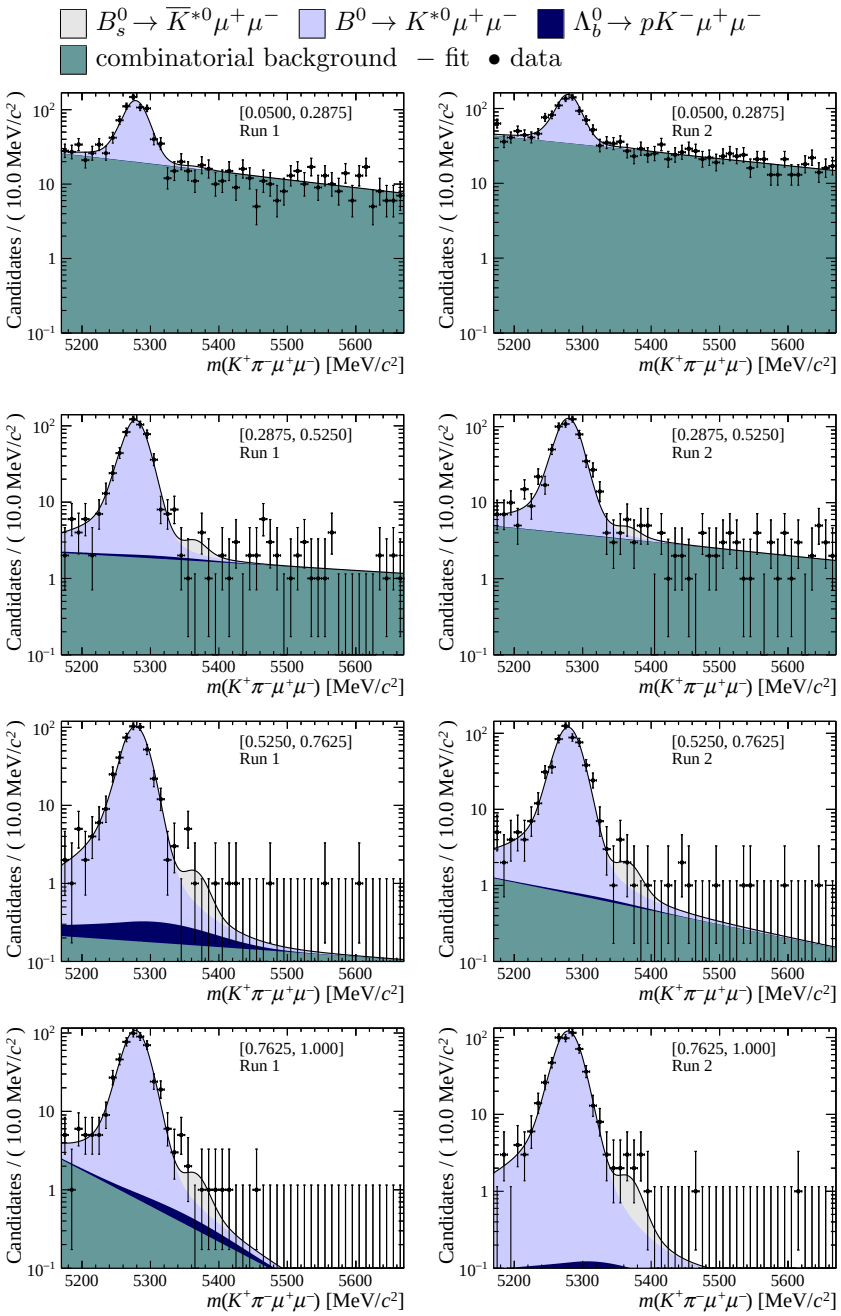

Figure 7.13: Distribution of reconstructed $K^+\pi^-\mu^+\mu^-$ invariant mass of candidates in the rare mode for the Run 1 (left) and Run 2 (right) data sets as well as the result of a maximum likelihood fit. The candidates are divided into four independent bins of increasing neural network response per data taking period.

To determine the significance of the $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ yield that has been obtained, Wilks' theorem is used. By taking the difference in the likelihood be-

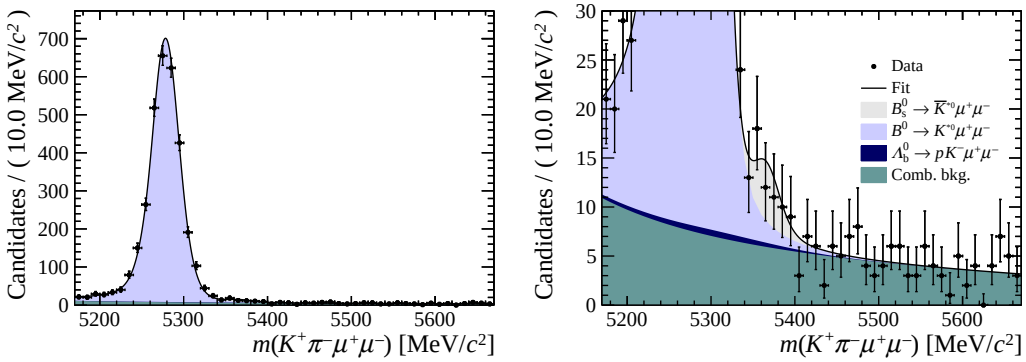

Figure 7.14: Distribution of reconstructed $K^+\pi^-\mu^+\mu^-$ invariant mass of candidates in the rare mode as well as the results of a maximum likelihood fit to these candidates. The three neural network bins of highest purity for each set of run conditions have been combined. The candidates are shown over the full range (left) and over a restricted vertical range to emphasise the $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ component (right). Fit components are detailed in the legend in the same order they are stacked in the plot.

tween the nominal fit described above and a fit under the null-hypothesis (that there is no $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ component), the significance can be determined as $\sqrt{-2\log(L_{S+B}/L_B)} = 3.4$, where L_{S+B} is the likelihood from the fit including the $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ component and L_B is the likelihood from the background only fit. The variation of $-2\log(L_{S+B}/L_B)$ with respect to the $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ yield is shown in Fig. 7.15. This log-likelihood variation and the resulting significance has been calculated including systematic uncertainties related to the raw yield of $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ decays, the details of which are further discussed in Sec. 8.1.

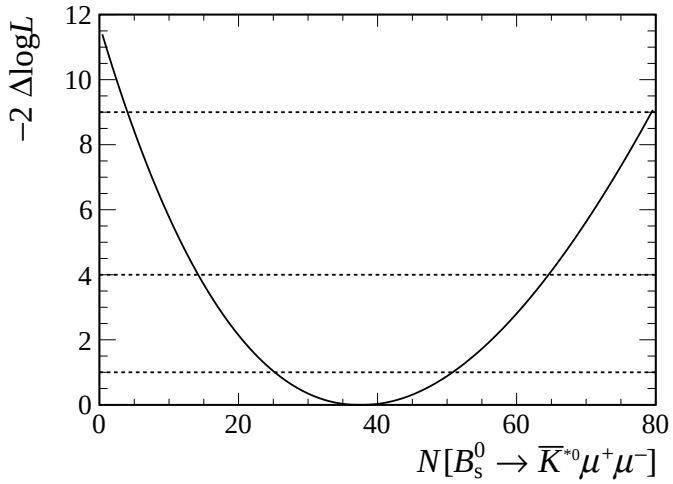

Figure 7.15: Change in log-likelihood with respect to the $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ yield from simultaneous fit to all bins. Systematic uncertainties on the yield have been included in the likelihood.

7.4 Branching fraction measurement

As initially discussed in Sec. 4.2, the yield of $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ decays that has been measured will be converted to a branching fraction via the relation:

$$\begin{aligned}\mathcal{B}(B_s^0 \to \overline{K}^{*0}\mu^+\mu^-) = \mathcal{B}(B^0 \to J/\psi K^{*0})\mathcal{B}(J/\psi \to \mu^+\mu^+) \\ \times \frac{f_d}{f_s}\frac{N(B_s^0 \to \overline{K}^{*0}\mu^+\mu^-)}{\varepsilon(B_s^0 \to \overline{K}^{*0}\mu^+\mu^-)}\frac{\varepsilon(B^0 \to J/\psi K^{*0})}{N(B^0 \to J/\psi K^{*0})}\,.\end{aligned} \tag{7.12}$$

The value of the external branching fractions that are input into this calculation have previously been given as well as the ratio of B^0 to B_s^0 production as measured by LHCb. The efficiency ratios are taken from Tab. 6.2, where the correlation between particle identification uncertainties has been taken into account. This results in an efficiency ratio of

$$\frac{\varepsilon(B^0 \to J/\psi K^{*0})}{\varepsilon(B_s^0 \to \overline{K}^{*0}\mu^+\mu^-)} = 1.669 \pm 0.022$$

to be used in the normalisation procedure. The main difference in the efficiencies comes from the relative difference in efficiency of the q^2 requirements between the two modes. The vetoes removing the J/ψ and $\psi(2S)$ regions in the rare mode reduce the efficiency of the selection, whereas the control mode does not have a similar reduction.

The final component to the normalisation procedure is a correction to the fitted yields to account for reconstructed $K^+\pi^-$ candidates which did not decay via a $K^{*0}(892)$ meson but instead via an S-wave configuration. For the $B^0 \to J/\psi K^{*0}$ mode, the S-wave fraction has previously been determined to be $F_{\rm S}(B^0 \to J/\psi K^{*0}) = (6.4 \pm 0.3 \pm 1.0)\%$ in the same $m(K^+\pi^-)$ window used in this analysis [131]. As this is the first search for the $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ mode, the S-wave fraction in the decay is unknown. In this case we make the assumption that the fraction will be similar to that for $B^0 \to K^{*0}\mu^+\mu^-$. By scaling the measurement of the S-wave from Ref. [46] to the $m(K^+\pi^-)$ window used in this analysis, the value of the S-wave fraction is found to be $F_{\rm S}(B^0 \to K^{*0}\mu^+\mu^-) = (3.4 \pm 0.8)\%$. Further details of how this is determined and the associated systematic uncertainty with the assumption that the contribution is similar between the two modes are given in Sec. 8.2.2.

Taking all the individual components described above, the measured branching fraction of the decay $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ is found to be

$$\mathcal{B}(B_s^0 \to \overline{K}^{*0}\mu^+\mu^-) = [2.9 \pm 1.0\,(\text{stat}) \pm 0.2\,(\text{syst}) \pm 0.3\,(\text{norm})] \times 10^{-8}\,,$$

where the first uncertainty is statistical, the second systematic (the calculation of which is discussed in Chapter 8) and the final due to external parameters in the normalisation procedure. This normalisation uncertainty constitutes the uncertainty on the external branching fractions, ratio of fragmentation fractions and the measured S-wave contributions.

In addition to the branching fraction measured with respect to the decay $B^0 \to J/\psi K^{*0}$, branching ratios with the decays $B^0 \to K^{*0}\mu^+\mu^-$ and $B_s^0 \to J/\psi \overline{K}^{*0}$ are also determined. The primary differences in calculating these ratios are the slightly different efficiency ratios of the $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ decay to these modes, different S-wave contributions and, in the case of the decay $B_s^0 \to J/\psi\overline{K}^{*0}$, no inclusion of a factor to account for the ratio of fragmentation fractions. For the ratio to $B^0 \to K^{*0}\mu^+\mu^-$, the efficiency ratio between the two modes is given by

$$\frac{\varepsilon(B^0 \to K^{*0}\mu^+\mu^-)}{\varepsilon(B_s^0 \to \overline{K}^{*0}\mu^+\mu^-)} = 0.937 \pm 0.014 \ ,$$

and the S-wave fractions are as previously determined. The resulting value of the branching fraction ratio is given by

$$\frac{\mathcal{B}(B_s^0 \to \overline{K}^{*0}\mu^+\mu^-)}{\mathcal{B}(B^0 \to K^{*0}\mu^+\mu^-)} = [3.3 \pm 1.1\,(\text{stat}) \pm 0.3\,(\text{syst}) \pm 0.2\,(\text{norm})] \times 10^{-2} \ .$$

For calculation the final branching ratio, the efficiency ratio between the $B_s^0 \to J/\psi\overline{K}^{*0}$ and $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ modes is given by

$$\frac{\varepsilon(B_s^0 \to J/\psi\overline{K}^{*0})}{\varepsilon(B_s^0 \to \overline{K}^{*0}\mu^+\mu^-)} = 1.792 \pm 0.025 \ .$$

The S-wave fraction is taken from the measurement provided in Ref [132] where the fraction is calculated in bins of $m(K^+\pi^-)$. This is then scaled to give an average S-wave contribution over the full $m(K^+\pi^-)$ range of $F_{\rm S}(B_s^0 \to J/\psi\overline{K}^{*0}) = (16.0 \pm 3.0)\%$. The measured value of the branching ratio is then given by

$$\frac{\mathcal{B}(B_s^0 \to \overline{K}^{*0}\mu^+\mu^-)}{\mathcal{B}(B_s^0 \to J/\psi\overline{K}^{*0})} = [1.4 \pm 0.4\,(\text{stat}) \pm 0.1\,(\text{syst}) \pm 0.1\,(\text{norm})] \times 10^{-2} \ .$$

Chapter 8

Systematic uncertainties

In the following chapter the sources of systematic uncertainty on the branching fraction measurement of $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ and details of how the values of these uncertainties are determined are described. The full summary of these uncertainties is given in Tab. 8.1, where the sources of uncertainty are given separately for each performed branching fraction measurement. The total systematic uncertainty on any measurement is taken as the sum in quadrature of the individual uncertainties. In addition to the systematic uncertainties, the size of the uncertainties related to the external parameters used in the normalisation are also shown.

As a result of the normalisation strategy used in this analysis, where all measurements are carried out as ratios of decay modes with the same four particle final state, the majority of systematic uncertainties cancel out, leaving relatively small contributions. In general, two classes of systematic uncertainties are considered: either an uncertainty that is associated to the determination of the rare $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ mode yield or as an uncertainty associated to the normalisation procedure used in the branching fraction determination. In the determination of the significance of the observed $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ yield described in Sec. 7.4 only the systematic uncertainties related to the yields are considered.

8.1 Uncertainties on the yield

8.1.1 Invariant mass lineshape distributions

The first set of uncertainties considered on the yield determination is that associated to the PDFs of the signal and background components. The PDF of the signal is determined from simulation and then the resolution and tail parameters are scaled by the result of the fit to the control mode. It is not guaranteed that this is a correct

description of the rare mode signal, as the resolution of the signal varies depending on the q^2 of the dimuon pair. This effect can be seen in the simulated samples and the resolution of the $B^0 \to K^{*0}\mu^+\mu^-$ candidates in bins of q^2 is shown in Fig. 8.1. To determine the size of this effect, the fit to the rare mode is repeated where the resolution of the signal is no longer fixed to the control mode but is instead Gaussian constrained to be within 0.5 MeV/c^2 of the result from the control mode, covering the variation of resolution over q^2 from the simulation. The difference between the nominal $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ yield and the yield from this alternative fit is then taken as the systematic uncertainty, an effect of 0.1%.

As the fit to the control mode can be carried out both with and without a constraint on the J/ψ mass, with significantly different line shapes for the signal in each case, a verification of the signal line shape can be carried out by comparing the results of the two fits. The ratio of branching fractions between the $B_s^0 \to J/\psi\overline{K}^{*0}$ and $B^0 \to J/\psi K^{*0}$ modes is used as a measure of the consistency between the two fits. The results of these fits are shown in Tab. 8.2, showing a 0.5% difference between the fits with and without the mass constraint which is taken as a systematic uncertainty on the signal line shape.

An uncertainty on the description of the $\Lambda_b^0 \to pK^-\mu^+\mu^-$ is determined by repeating the fit to the rare mode with an alternate PDF describing the component.

Table 8.1: Main sources of systematic uncertainty considered on the branching fraction measurements. The first uncertainty applies to the measurement of $\mathcal{B}(B_s^0 \to \overline{K}^{*0}\mu^+\mu^-)$, the second to $\mathcal{B}(B_s^0 \to \overline{K}^{*0}\mu^+\mu^-)/\mathcal{B}(B^0 \to K^{*0}\mu^+\mu^-)$ and the third to $\mathcal{B}(B_s^0 \to \overline{K}^{*0}\mu^+\mu^-)/\mathcal{B}(B_s^0 \to J/\psi\overline{K}^{*0})$, respectively. A description of the different contributions can be found in the text. The first three sources of uncertainty affect the measured yield of the signal decay. The total uncertainty is the sum in quadrature of the individual sources. The final row indicates the additional uncertainty arising from the uncertainties on external parameters used in the measurements.

Source	Uncertainties		
	$\mathcal{B}(B_s^0 \to \overline{K}^{*0}\mu^+\mu^-)$	$\frac{\mathcal{B}(B_s^0 \to \overline{K}^{*0}\mu^+\mu^-)}{\mathcal{B}(B^0 \to K^{*0}\mu^+\mu^-)}$	$\frac{\mathcal{B}(B_s^0 \to \overline{K}^{*0}\mu^+\mu^-)}{\mathcal{B}(B_s^0 \to J/\psi\overline{K}^{*0})}$
Mass lineshapes	0.5%	0.5%	0.5%
Neural network response	0.5%	0.5%	0.5%
Residual background	2.0%	2.0%	2.0%
Decay models	4.0%	4.0%	4.0%
Non-$\overline{K}^{*0}$ states	3.4%	3.4%	3.4%
Efficiency	1.3%	1.5%	1.4%
Data-simulation differences	2.2%	2.2%	0.8%
Total systematic uncertainty	6.2%	6.3%	5.9%
External parameters	8.9%	5.9%	4.0%

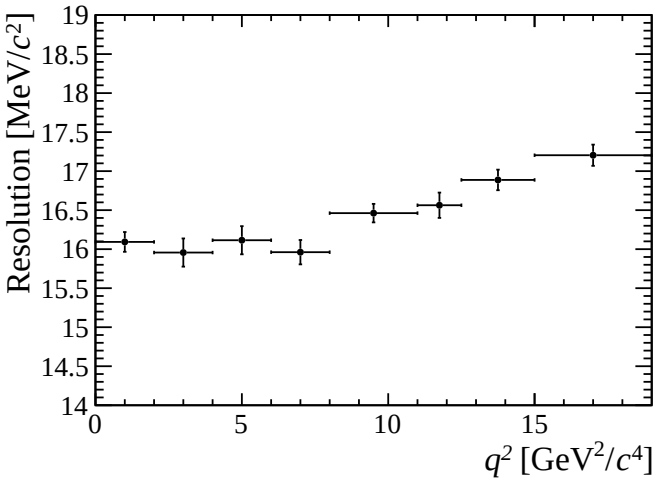

Figure 8.1: Resolution on the reconstructed $K^+\pi^-\mu^+\mu^-$ invariant mass of simulated $B^0 \to K^{*0}\mu^+\mu^-$ candidates in bins of dimuon invariant mass squared (q^2).

Table 8.2: Measured ratio of branching fractions of $B_s^0 \to J/\psi\overline{K}^{*0}$ to $B^0 \to J/\psi K^{*0}$ candidates for both control mode fits.

	Standard fit	Constrained fit
$N_{B_s^0}/N_{B^0}$	$(9.31 \pm 0.24) \times 10^{-3}$	$(9.14 \pm 0.15) \times 10^{-3}$
$\epsilon_{B^0}/\epsilon_{B_s^0}$	0.919 ± 0.009	0.932 ± 0.009
$\mathcal{B}(B_s^0 \to J/\psi\overline{K}^{*0})/\mathcal{B}(B^0 \to J/\psi K^{*0})$	$(3.30 \pm 0.21) \times 10^{-2}$	$(3.29 \pm 0.20) \times 10^{-2}$

Instead of the Crystal-Ball line shape used in the standard fit an alternate model of a gaussian core with an exponential tail is used. The systematic uncertainty is then taken as the difference in rare mode yield between the two fits, resulting in an uncertainty of 0.1%. The total systematic uncertainty associated to mass line shapes is taken as the sum in quadrature of those described above, leading to a total uncertainty of 0.5%.

8.1.2 Neural network binning variations

The fit to the rare mode is carried out under the assumption that the signal efficiency is equal in each bin of neural network response and that the efficiency is equal for both the $B^0 \to K^{*0}\mu^+\mu^-$ mode and $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ mode. In Fig. 5.23 it can be seen that this assumption is true within statistical uncertainties as determined from simulation. There are two aspects of this assumption that should be tested: first, that a variation within the statistical uncertainty of the efficiency from MC does

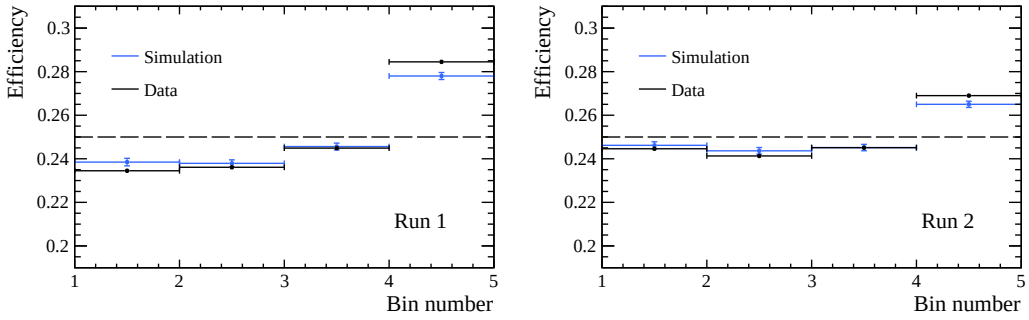

Figure 8.2: Efficiencies in bins of neural network output for the $B^0 \to J/\psi K^{*0}$ control mode from both simulation and a fit to data for Run 1 (left) and Run 2 (right). The black horizontal line indicates equal efficiency across all bins.

not affect the final result and second, that the simulated efficiency matches the true efficiency from real data.

To test the first aspect, the fit to the rare mode is carried out without the efficiencies of the $B^0 \to K^{*0}\mu^+\mu^-$ and $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ modes fixed to be exactly 1/4 in each bin but instead Gaussian constrained to the signal efficiency as determined from simulated with a width equal to statistical uncertainty due to the size of the simulated samples. This allows for a variation of the efficiency within this statistical uncertainty and for the two modes to have slightly different efficiencies. The difference in the yield of the $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ mode in this fit and the standard fit is then taken as a systematic uncertainty, an effect of 0.3%.

The second aspect is tested by making use of the $B^0 \to J/\psi K^{*0}$ control mode and comparing signal efficiency in bins of neural network response to that from simulation. In the fit to the control mode, the fraction of the signal in each bin is a floating parameter and so can be compared to the fraction from simulation. Figure 8.2 shows a comparison of the binned signal efficiency from both simulation and the fit to the control mode. It is clear here that the differences between the two are small although not completely consistent within statistical uncertainties. The effect of this residual difference is tested by repeating the fit and correcting the efficiencies in each bin by the absolute difference between simulation and data taken from the control mode, chosen as the extreme situation. Carrying this out there is a 0.4% difference in $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ yield between the the nominal fit and the investigated one and this is assigned as the systematic uncertainty. The total systematic uncertainty related to neural network binning is then taken as 0.5%.

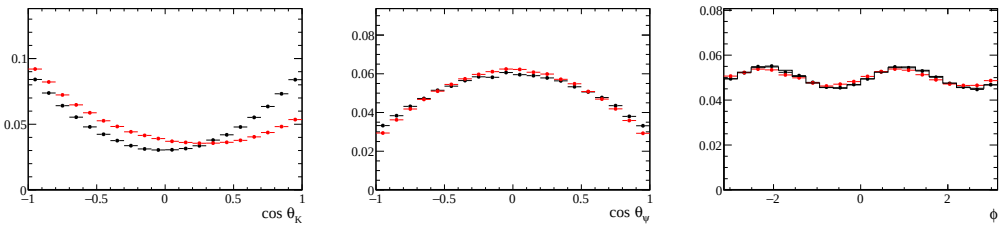

Figure 8.3: Angular distributions (from left to right) for the cosine of the angle between the kaon and pion, the cosine of the angle between the two muons and the angle between the two planes defined by the two sets of daughter particles for $B^0 \to J/\psi K^{*0}$ decays. Points in black are from MC, points in red from a model based on the measurement by Belle.

8.1.3 Residual backgrounds

The final uncertainty related to the $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ yield is one from residual backgrounds in the signal region. As calculated in Sec. 5.2.1, there are potential residual backgrounds from the decay $B_s^0 \to K^{*0}(892)K^{*0}(892)$ which could contribute up the 2% of the measured signal yield with a negligible effect on the $B^0 \to K^{*0}\mu^+\mu^-$ yield. This is included directly as an uncertainty on the $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ yield. In calculating the significance of the observed yield including this systematic uncertainty, a simple gaussian component centred in the signal region with a yield gaussian constrained to the calculated yield for the background with the associated uncertainty is included in the fit model.

8.2 Normalisation uncertainties

8.2.1 Simulated decay models

The primary systematic uncertainty in the branching fraction calculation (excluding uncertainties due to external parameters) is due to differences in the angular distribution of the decays between simulation and taken from data. A difference in the angular distribution can significantly affect the efficiency to reconstruct the decay mode and thus change the result of the normalisation. To demonstrate this difference, Fig. 8.3 shows the angular distribution of $B^0 \to J/\psi K^{*0}$ decays taken from both the MC simulation used in this analysis and from measurements carried out by the Belle collaboration [116]. The difference in the two is clear in the cosine of the angle between the kaon and the pion, $\cos(\theta_K)$.

To evaluate the effect of this difference, the ratio of the measured angular distribution and that from simulation is applied as a weight to correct the simulated distribution. The difference between the corrected and uncorrected efficiency ratios

is then taken as the systematic uncertainty. A similar procedure is carried out for the $B_s^0 \to J/\psi \overline{K}^{*0}$ mode, correcting the simulated distribution to match that measured by LHCb [133] and once again applying weights to correct the reconstruction efficiency. Finally, an uncertainty on the $B^0 \to K^{*0}\mu^+\mu^-$ is determined by weighting the simulated sample such that the Wilson coefficients in the decay model used are consistent with the values from global fits to $b \to s$ data, previously described in Sec. 2.3.6. The total effects of these model variations lead to a 4% uncertainty on the branching fraction normalisation for each ratio of measurements.

8.2.2 Non-$\overline{K}^{*0}$ states

As discussed in Sec. 7.4, the contribution from non resonant $K^+\pi^-$ contributions in the $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ mode is unknown. In the correction to the $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ yield, it is assumed that the S-wave contribution is the same as it is in $B^0 \to K^{*0}\mu^+\mu^-$. The $B^0 \to K^{*0}\mu^+\mu^-$ S-wave fraction is taken from the LHCb measurement [46], a measurement that is carried out in a $100\,\text{MeV}/c^2$ window around the $K^{*0}(892)$ mass. To get the correct S-wave fraction to use in this analysis (a $70\,\text{MeV}/c^2$ window), this fraction is scaled making use of the S-wave model in terms of $m(K^+\pi^-)$. This fraction as a function of q^2 is shown in Fig. 8.4 and leads to the average S-wave fraction of 3.4% that is used in the corrections to the yields. As it is only assumed that this fraction is similar for the $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ mode, a conservative uncertainty of 100% of this fraction is applied, leading to the 3.4% systematic uncertainty related to this unknown contribution.

8.2.3 Efficiency calculation

The remaining systematics associated to the branching fraction normalisations results from the calculation of efficiency ratios from simulation. The first of these uncertainties results from the limited sample sizes of the MC simulation, leading to a statistical uncertainty on the derived efficiencies. As the simulated sample size for each decay mode is slightly different, the relative uncertainty also differs. This leads to slightly different uncertainties on the efficiency ratios used in each of the branching fraction/ratio calculations, as displayed in Tab. 8.1. As detailed in Sec. 6.3, the correlations of the PID selection efficiencies between decay modes are taken into account in the calculation of the efficiency ratios, resulting in a lower systematic uncertainty then would be expected from the raw efficiencies per decay mode.

The final uncertainty resulting from the efficiency calculations derives from the corrections applied to simulation. The need to correct the simulated samples

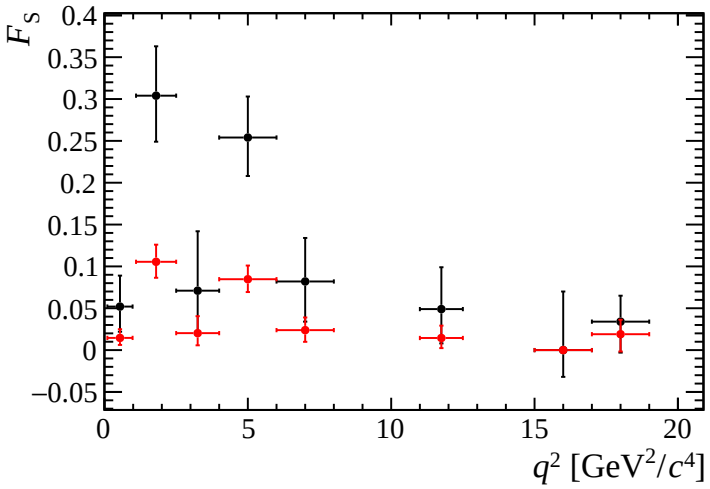

Figure 8.4: S-wave fraction in bins of q^2 for the wide $m(K\pi)$ window of the $B^0 \to K^{*0}\mu^+\mu^-$ S-wave fraction measurement (in black) and scaled according to the $m(K\pi)$ model into the $m(K\pi)$ range used in this analysis.

shows that there is an inherent uncertainty in the efficiencies taken from simulation. However, the differences in efficiency ratios calculated with and without corrections are small and this is one of the primary motivations of carrying out these measurements as ratios. A measure of the residual uncertainty from these efficiencies is taken as the extreme case of the difference in efficiency ratios between corrected and uncorrected simulation. The total uncertainty, taking into account corrections due to the kinematic weighting described in Sec. 5.1.3 and the tracking efficiency and muon identification corrections described in Sec. 6.4, is 0.8%. In addition, when the measurement is performed between a B_s^0 and a B^0 decay mode, the extra correction applied to correct the B_s^0 production kinematics leads to an additional residual difference and after adding the uncertainty in quadrature brings the total to 2.2%.

Chapter 9

Summary

A first search for the rare decay mode $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$, making use of a total of $4.6\,\text{fb}^{-1}$ of data collected by the LHCb experiment over both LHC Run 1 and Run 2, has been presented. The full LHCb Run 1 and approximately one third of the Run 2 has been used as part of this search. The first evidence for this decay mode has been found, with a yield of 38 ± 12 $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ decays, giving a significance of 3.4 standard deviations above the null hypothesis. By measuring this decay mode relative to the decay $B^0 \to J/\psi K^{*0}$, the first measurement of its branching fraction has been made and found to be

$$\mathcal{B}(B_s^0 \to \overline{K}^{*0}\mu^+\mu^-) = [2.9 \pm 1.0\,(\text{stat}) \pm 0.2\,(\text{syst}) \pm 0.3\,(\text{norm})] \times 10^{-8}\ .$$

In addition, measurements of the branching fraction relative to two other modes have been made with the ratio with respect to the decay $B^0 \to K^{*0}\mu^+\mu^-$ measured to be

$$\frac{\mathcal{B}(B_s^0 \to \overline{K}^{*0}\mu^+\mu^-)}{\mathcal{B}(B^0 \to K^{*0}\mu^+\mu^-)} = [3.3 \pm 1.1\,(\text{stat}) \pm 0.3\,(\text{syst}) \pm 0.2\,(\text{norm})] \times 10^{-2}\ ,$$

and the ratio with respect to the decay $B_s^0 \to J/\psi\overline{K}^{*0}$ measured to be

$$\frac{\mathcal{B}(B_s^0 \to \overline{K}^{*0}\mu^+\mu^-)}{\mathcal{B}(B_s^0 \to J/\psi\overline{K}^{*0})} = [1.4 \pm 0.4\,(\text{stat}) \pm 0.1\,(\text{syst}) \pm 0.1\,(\text{norm})] \times 10^{-2}\ .$$

These measurements are consistent with predictions of the SM branching fraction of this mode presented in Refs. [112–114], where the most recent prediction of the branching fraction is $(3.4 \pm 0.8) \times 10^{-8}$. The ratio with respect to the mode $B^0 \to K^{*0}\mu^+\mu^-$ is also consistent with a naive scaling between these two modes of the square of the CKM factor $|V_{td}/V_{ts}|$. Further input from the theory community

to calculate the hadronic form factors of these two decay modes along with the correlations between the form factors will be necessary to convert the branching ratio into a direct measurement of $|V_{td}/V_{ts}|$.

The heavy suppression of this decay mode coupled with the presence of a B^0 partner to the mode which occurs at a rate approximately 100 times greater has meant that great care has been needed in choosing the strategy to search for this mode. The detailed studies of the backgrounds involved in the analysis as well as the techniques used to optimise selection criteria and provide strong constraints on any residual background components have been essential in making this measurement possible and will set the ground work for future studies of this mode and similar decay modes.

The measurement presented here is an important first step towards the further study of the rare $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ decay. In the near future, the data collected by the LHCb experiment up to the end of Run 2 of the LHC will provide the opportunity for the first observation of this decay mode as well as an improvement in the precision of any measurement of $|V_{td}/V_{ts}|$ that may be extracted from the combination of the decays $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ and $B^0 \to K^{*0}\mu^+\mu^-$. This decay mode is also a key channel in future searches for new physics making use of $b \to d\ell^+\ell^-$ transitions, with a potential angular analysis of this decay to a similar precision as the decay $B^0 \to K^{*0}\mu^+\mu^-$ providing part of the physics case for a future upgrade of the LHCb experiment [108]. In particular, the $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$ mode is one of the best prospects for carrying out an angular analysis of a $b \to d\ell^+\ell^-$ transition, as the decay is self tagging, unlike $B^0 \to \pi^+\pi^-\mu^+\mu^-$, and has a well studied intermediate resonance structure, unlike $\Lambda_b^0 \to p\pi^-\mu^+\mu^-$. A full angular analysis will be possible with well defined theoretical predictions for the angular observables. Measurements of this kind will be vital to probe the flavour structure of any new physics scenario that is presented as a result of the current flavour anomalies.

Appendix A

Upgrade VELO design studies

As discussed in Sec. 3.3, an upgrade of the LHCb detector will be carried out during the long shut down period between Run 2 and Run 3 of the LHC. As part of this upgrade process, the VELO sub-detector will be replaced with a new silicon pixel based detector. Many design choices for this sub-detector have had to be made, using simulated events to evaluate the performance of design proposals. In this appendix, details of two studies carried out as part of this decision making process are given, the first a study to choose a support structure for the new VELO modules and the second a study to choose a cooling system for the modules.

To perform these studies, the design schematics of the proposals need to be converted into a format that can be read by the LHCb simulation framework. The LHCb detector description database contains information on the geometry and materials of all the components of the detector. These individual components are stored in XML (eXtendable Markup Language) files, where the geometry of the components is described in the form of boolean union, subtractions and intersections of simple geometrical shapes, *e.g.* boxes or tubes. By describing the design schematics in this form, new detector descriptions can be added to the detector description database and simulated samples can be generated with the design proposals in place.

The work described in this appendix was carried out in collaboration with other members of the LHCb VELO group, and has previously been described in LHCb internal documents, where Ref. [134] details the support design choice and Ref. [135] details the cooling design choice. The specific contribution to these studies made by the author was to convert the design schematics in the XML format and then validate that these are modelled correctly within the simulation.

A.1 Module support design

The upgraded VELO will consist of 26 pairs of modules arranged around the interaction point of the LHCb detector, in much the same way as the current VELO sub-detector. Each of these modules needs to mounted on a support structure which connects the modules to the retractable ends of the VELO sub-detector. The design requirements of these supports are that they are structurally capable of holding the modules in place during detector operations, while limiting the material that is near the interaction point of the experiment. An increased material budget increases the likelihood that a track will undergo secondary scattering within the detector, reducing the resolution of any measurement of the tracks that can be made. There were two competing designs for the module support and they are henceforth referred to as the "Oxford" design and the "Nikhef" design, named after the institutes which proposed each of the designs.

A.1.1 Oxford design

The Oxford design consists of a single sheet of carbon fibre upon which the module is mounted. Two stainless steel coolant delivery pipes are then connected to the substrate of the module via an Invar manifold. The data connectors of the module are raised up slightly from the module and placed upon a secondary piece of carbon fibre on the support sheet. The full Oxford design for the support system is shown in Fig. A.1, where the support sheet, cooling pipes and the VELO module are all shown. Note, that due to limitations in how logical volumes can be written in the XML format, the cooling pipes shown are not the same shape as in the true design. In the true design, the joints between the pipe sections are curved rather than at right angles. The dimensions of the straight pipe elements are such that the same amount of material is used as there would be if they were curved pipe elements.

A.1.2 Nikhef design

The Nikhef support design consists of two cylindrical carbon fibre legs connected by a carbon fibre mid-plate upon which the VELO module is mounted. Two smaller carbon fibre plate sections hold in place an Invar manifold that connects two stainless steel cooling pipes to the module substrate. The full Nikhef design for the support system is shown in Fig. A.2, where the carbon fibre support structure, cooling pipes and the VELO module are all shown.

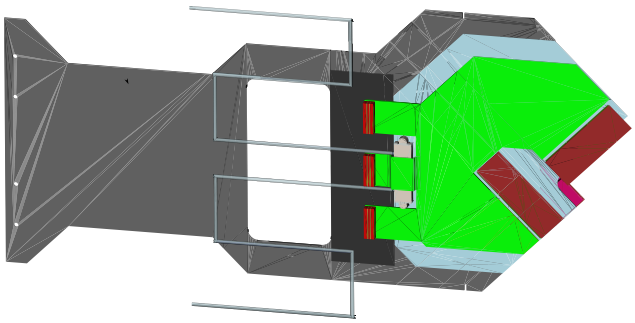

Figure A.1: Design of the Oxford support proposal where a module has been mounted on the support, as modelled in the LHCb detector description database. Different colours in the figure indicate different materials, where the key materials for the support structure are carbon fibre (shown in two shades of dark grey), stainless steel (shown in silver) and Invar (light grey).

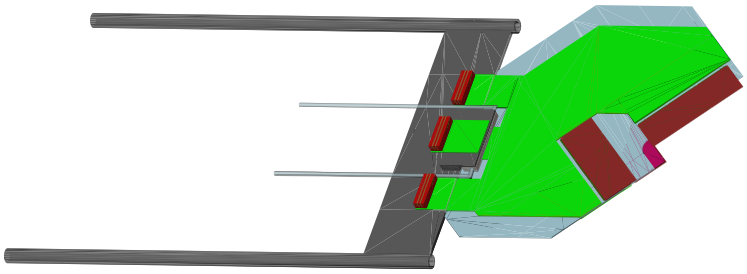

Figure A.2: Design of the Nikhef support proposal where a module has been mounted on the support, as modelled in the LHCb detector description database. Different colours in the figure indicate different materials, where the key materials for the support structure are carbon fibre (dark grey), stainless steel (shown in silver) and Invar (light grey).

A.1.3 Simulated performance

The total material of the two support designs can be evaluated with a ray-tracing method using the XML description of the designs shown above. This material is shown in terms of thickness in the z direction of the LHCb coordinate system (in the direction of the beam line), both in mm and as a fractional radiation length of the particles moving through the material. The radiation length is based on the amount and the density of material that a particle passes through. These material measurements are shown for both designs in the $x - y$ plane of the detector in Fig. A.3 and the differences between the two designs are clearly visible. There is, however, very little difference between the designs near the interaction point at the centre of these figures. This is the key region for detector performance and as such

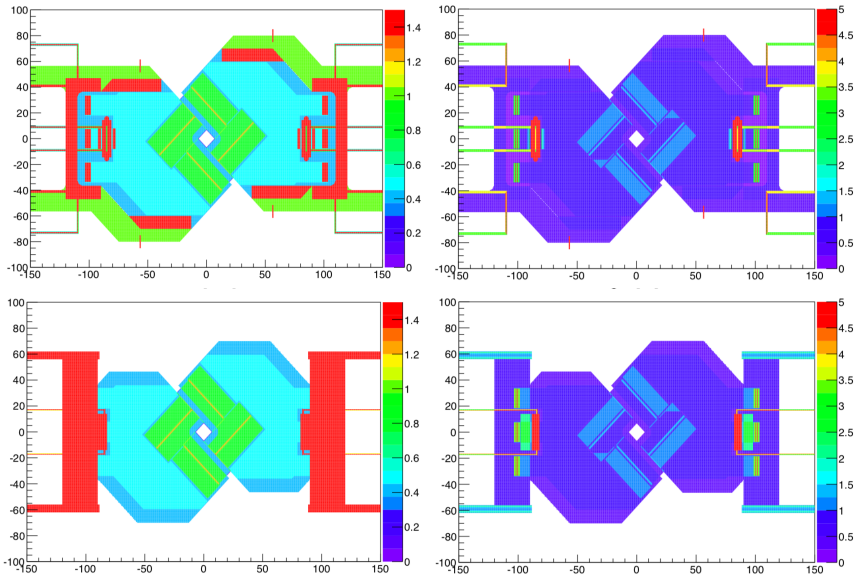

Figure A.3: Material thickness in the z (beam) direction of a single pair of VELO modules in both mm (left) and fractional number of radiation lengths (right). This thickness is shown for the case where the modules are mounted on the Oxford support design (top) and the Nikhef support design (bottom). The x and y axis correspond to the x and y direction in the LHCb detector in dimensions of mm.

it is expected that the two designs will have similar performance.

The performance of the two designs can be further investigated by making use of simulated decay samples. Samples are generated and then passed through the full simulation chain for both of the proposed designs. These simulated candidates are then reconstructed for both detector descriptions and the resolution of the impact parameter can be determined using the knowledge of the true (generated) track position and the reconstructed track position. The track finding efficiency as a function of both pseudorapidity and azimuthal angle in the detector is also determined. The results of these performance studies are shown in Fig. A.4 and show that the performance of the two designs are consistent with each other within the uncertainties of the measurements. As such, there is no preference for a given design in terms of physics performance and the decision on which design to use is instead made in terms of the practicality of the manufacturing process.

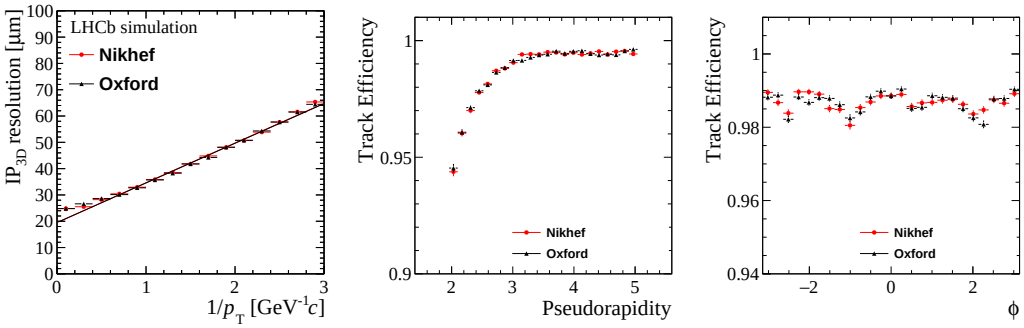

Figure A.4: Track 3-dimensional IP resolution as a function of $(1/p_T)$ (left) and track finding efficiency as function of pseudorapidity (middle) and azimuthal angle (right) as determined from simulated samples for both the Oxford support design and the Nikhef support design.

A.2 Module cooling design

Two module cooling proposals have been presented for the upgraded VELO, both designed to match the requirement that the coolant is delivered directly onto the module to deal with high temperatures from the module read-out electronics and the radiation near the interaction point of the detector. The first design makes use of a silicon substrate that has had approximately 100 µm deep micro-channels bored into the surface that allow coolant to flow, referred to as the Plan A design. The second uses an aluminium nitride substrate with stainless steel capillary pipes running through it to deliver the coolant, referred to as the Plan B design. In both cases the module sensors and read-out electronics of the module are mounted directly onto the cooled substrate.

A.2.1 Plan A

The Plan A cooling design consists of a 0.5 mm thick sheet of silicon as a substrate of the module which is connected to two stainless steel coolant delivery pipes via an Invar manifold system at the end of the substrate furthest from the interaction point. The density of the silicon material used is adjusted to account for the presence of the cooling micro-channels in the substrate. The full Plan A cooling design is shown in Fig. A.5, where the cooling system, a support structure based on the Nikhef design described above and the rest of the VELO sensor module are included.

A.2.2 Plan B

The Plan B cooling design consists of a 1 mm thick sheet of aluminium nitride which has had 0.7 mm thick channels cut out to allow a series of coolant capillary pipes to run through the substrate. Four of these stainless steel capillary pipes run through the substrate, where in this geometry description curved sections of the pipes are modelled as right angles in the same way as described in Sec. A.1.1. Each capillary runs through the entire substrate before exiting, leaving eight capillary pipe ends that connect to a stainless steel manifold away from the VELO modules. This module then connects to two stainless steel coolant delivery pipes. The full Plan B cooling design is shown in Fig. A.6, where the cooling system, a support structure based on the Nikhef design described above and the rest of the VELO sensor module are included.

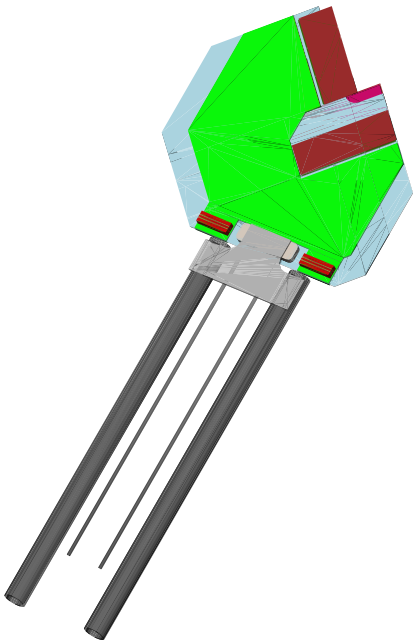

Figure A.5: Design of the Plan A cooling system for the VELO as part of a single module mounted on a support, as modelled in the LHCb detector description database. The silicon substrate through which the micro-channel cooling system runs is shown in light blue, the coolant delivery pipes are shown in dark grey and the cooling manifold block in shown in light grey.

A.2.3 Simulated performance

In the same manner as is done for the two support designs, the total material of the two cooling designs can be evaluated by using a ray-tracing method. Once again, the material is shown in terms of thickness in mm and in terms of a fractional radiation length. These material measurements are shown for both designs in the $x - y$ plane of the detector in Fig. A.7. For the cooling designs, the difference in material budget between the two proposals is significantly larger then for the support designs. Most importantly, the radiation length in the region close to the interaction point is significantly greater for the Plan B proposal, mainly driven by the thicker substrate that is required to house the cooling capillaries and the presence of the cooling capillaries themselves. Given this large difference, it is expected that there will be a noticeable difference in the physics performance of the two designs.

The IP resolution and track efficiency as a function of pseudorapidity and azimuthal angle in the detector are shown for both Plan A and Plan B in Fig. A.8. Also shown in these plots are the resolution and track efficiency when evaluated using a detector description previously used in studies of the upgraded VELO in the Upgrade VELO Technical Design Report (TDR) [103], used as a baseline for these

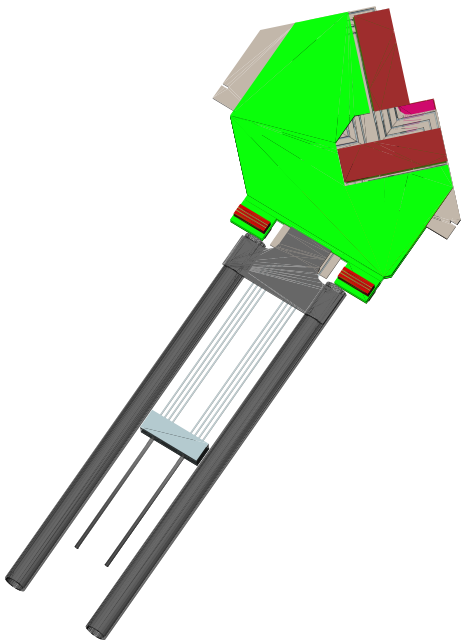

Figure A.6: Design of the Plan B cooling system for the VELO as part of a single module mounted on a support, as modelled in the LHCb detector description database. The aluminium nitride substrate is shown in light grey and the stainless steel cooling capillaries and manifold are shown in silver.

new studies. While the cooling system envisioned in the TDR also uses a micro-channel cooling system, the geometry and design of the VELO system has evolved significantly since the description given in the TDR and so it is not expected that the results from Plan A will be the same as those for the TDR description. It is clear from these plots that the performance of the Plan A solution is better than the Plan B solution, as expected. Barring any issues in the manufacturing procedure of the micro-channel cooling in the substrate, Plan A is the clear preferred cooling solution.

A.3 Summary

To aid in the choice between proposed designs for both the support structure and the cooling system of the upgraded LHCb VELO, the design schematics for these proposals have been converted into an XML format that can then be used in the detector description database for the LHCb upgrade simulation framework. Simulated samples have been produced using these descriptions which allow the physics performances of the designs to be evaluated.

For the two presented support designs, no significant different in physics

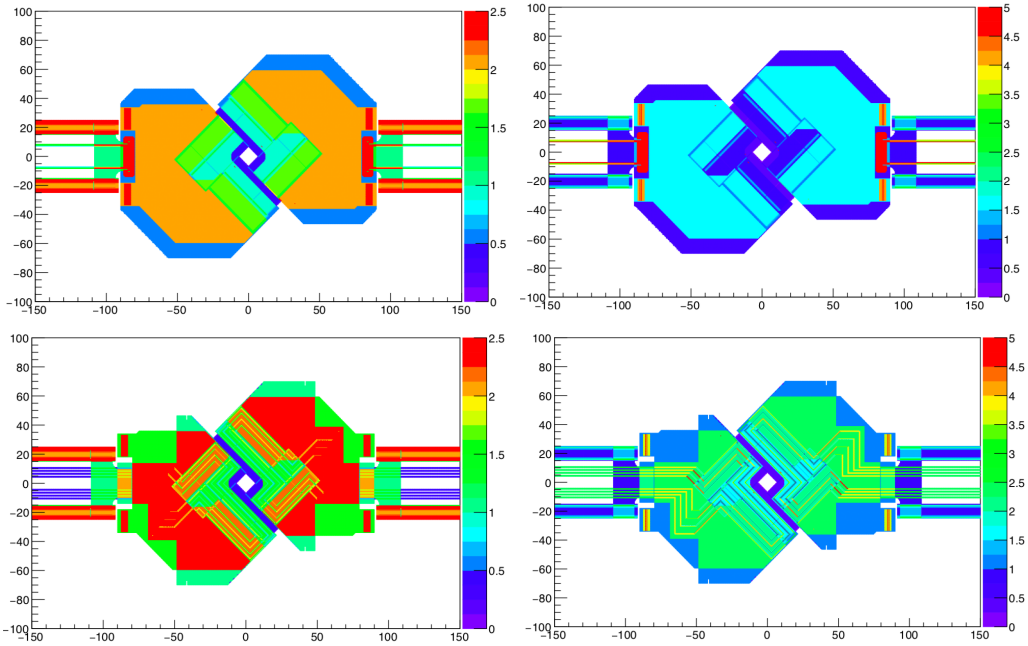

Figure A.7: Material thickness in the z (beam) direction of a single pair of VELO modules in both mm (left) and fractional number of radiation lengths (right). This thickness for both the Plan A (top) and Plan B (bottom) cooling design proposals are shown. The x and y axis correspond to the x and y direction in the LHCb detector in dimensions of mm.

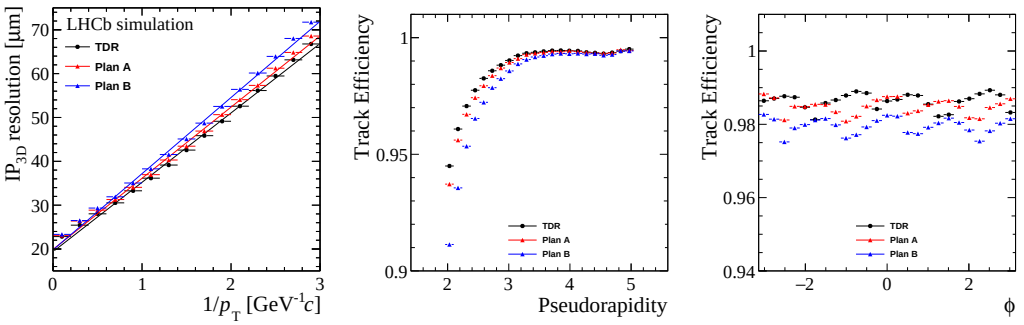

Figure A.8: Track 3-dimensional IP resolution as a function of $(1/p_T)$ (left) and track finding efficiency as function of pseudorapidity (middle) and azimuthal angle (right) as determined from simulated samples for the Plan A cooling design, the Plan B cooling design and the cooling design used in the upgrade VELO TDR.

performance is observed and the final decision on the design was made based on other, manufacturing based criteria that are beyond the scope of this work. The final design decision was in favour of the Nikhef design.

For the proposed cooling designs, the Plan B design introduces a significant amount of extra material into the acceptance of the LHCb detector with respect to the Plan A design and this leads to a visible decrease in physics performance. This

information on the physics performance of the two designs played a key part in the decision making process of the design choice, resulting in Plan A being chosen.

Bibliography

[1] LHCb collaboration, R. Aaij *et al.*, "Evidence for the decay $B_s^0 \to \overline{K}^{*0}\mu^+\mu^-$", *JHEP* **07** (2018) 020, arXiv:1804.07167.

[2] Particle Data Group, C. Patrignani *et al.*, "Review of particle physics", *Chin. Phys.* **C40** (2016) 100001, and 2017 update.

[3] F. Halzen and A. D. Martin, *Quarks and Leptons: An Introductory Course in Modern Particle Physics*, Wiley, New York, USA, 1984.

[4] S. L. Glashow, "Partial Symmetries of Weak Interactions", *Nucl. Phys.* **22** (1961) 579.

[5] S. Weinberg, "A Model of Leptons", *Phys. Rev. Lett.* **19** (1967) 1264.

[6] A. Salam, "Weak and Electromagnetic Interactions", *Conf. Proc.* **C680519** (1968) 367.

[7] S. L. Glashow, J. Iliopoulos, and L. Maiani, "Weak Interactions with Lepton-Hadron Symmetry", *Phys. Rev.* **D2** (1970) 1285.

[8] Wikimedia Commons, "Standard Model of Elementary Particles", https://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg, Accessed: 2018-09-16, Public domain.

[9] Belle collaboration, R. Mizuk *et al.*, "Dalitz analysis of $B \to K\pi^+\psi'$ decays and the $Z(4430)^+$", *Phys. Rev.* **D80** (2009) 031104, arXiv:0905.2869.

[10] LHCb collaboration, R. Aaij *et al.*, "Observation of the resonant character of the $Z(4430)^-$ state", *Phys. Rev. Lett.* **112** (2014) 222002, arXiv:1404.1903.

[11] BESIII, M. Ablikim *et al.*, "Observation of a Charged Charmoniumlike Structure in $e^+e^- \to \pi^+\pi^- J/\psi$ at $\sqrt{s} = 4.26\,\text{GeV}$", *Phys. Rev. Lett.* **110** (2013) 252001, arXiv:1303.5949.

[12] Belle collaboration, Z. Q. Liu *et al.*, "Study of $e^+e^- \to \pi^+\pi^- J/\psi$ and Observation of a Charged Charmoniumlike State at Belle", *Phys. Rev. Lett.* **110** (2013) 252002, arXiv:1304.0121.

[13] LHCb collaboration, R. Aaij *et al.*, "Observation of $J/\psi\phi$ structures consistent with exotic states from amplitude analysis of $B^+ \to J/\psi\phi K^+$ decays", *Phys. Rev. Lett.* **118** (2017) 022003, arXiv:1606.07895.

[14] LHCb collaboration, R. Aaij *et al.*, "Observation of $J/\psi p$ Resonances Consistent with Pentaquark States in $\Lambda_b^0 \to J/\psi K^- p$ Decays", *Phys. Rev. Lett.* **115** (2015) 072001, arXiv:1507.03414.

[15] F. Englert and R. Brout, "Broken Symmetry and the Mass of Gauge Vector Mesons", *Phys. Rev. Lett.* **13** (1964) 321.

[16] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons", *Phys. Rev. Lett.* **13** (1964) 508.

[17] P. W. Higgs, "Spontaneous Symmetry Breakdown without Massless Bosons", *Phys. Rev.* **145** (1966) 1156.

[18] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, "Global Conservation Laws and Massless Particles", *Phys. Rev. Lett.* **13** (1964) 585.

[19] ATLAS collaboration, G. Aad *et al.*, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Phys. Lett.* **B716** (2012) 1, arXiv:1207.7214.

[20] CMS collaboration, S. Chatrchyan *et al.*, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", *Phys. Lett.* **B716** (2012) 30, arXiv:1207.7235.

[21] N. Cabibbo, "Unitary Symmetry and Leptonic Decays", *Phys. Rev. Lett.* **10** (1963) 531.

[22] M. Kobayashi and T. Maskawa, "CP Violation in the Renormalizable Theory of Weak Interaction", *Prog. Theor. Phys.* **49** (1973) 652.

[23] L. Wolfenstein, "Parametrization of the Kobayashi-Maskawa Matrix", *Phys. Rev. Lett.* **51** (1983) 1945.

[24] A. D. Sakharov, "Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe", *Pisma Zh. Eksp. Teor. Fiz.* **5** (1967) 32, [Usp. Fiz. Nauk161,no.5,61(1991)].

[25] CKMfitter Group, J. Charles *et al.*, "CP violation and the CKM matrix: Assessing the impact of the asymmetric B factories", *Eur. Phys. J.* **C41** (2005) 1, arXiv:hep-ph/0406184.

[26] Y. Sofue and V. Rubin, "Rotation curves of spiral galaxies", *Ann. Rev. Astron. Astrophys.* **39** (2001) 137, arXiv:astro-ph/0010594.

[27] R. Massey, T. Kitching, and J. Richard, "The dark matter of gravitational lensing", *Rept. Prog. Phys.* **73** (2010) 086901, arXiv:1001.1739.

[28] L. Canetti, M. Drewes, and M. Shaposhnikov, "Matter and Antimatter in the Universe", *New J. Phys.* **14** (2012) 095012, arXiv:1204.4186.

[29] S. P. Martin, "A Supersymmetry primer", *Adv. Ser. Direc. High Energy Phys.* **21** (2010) 1, arXiv:hep-ph/9709356.

[30] R. P. Feynman and M. Gell-Mann, "Theory of Fermi interaction", *Phys. Rev.* **109** (1958) 193.

[31] K. G. Wilson and W. Zimmermann, "Operator product expansions and composite field operators in the general framework of quantum field theory", *Commun. Math. Phys.* **24** (1972) 87.

[32] T. Blake, G. Lanfranchi, and D. M. Straub, "Rare B Decays as Tests of the Standard Model", *Prog. Part. Nucl. Phys.* **92** (2017) 50, arXiv:1606.00916.

[33] S. Descotes-Genon, J. Matias, M. Ramon, and J. Virto, "Implications from clean observables for the binned analysis of $B \to K^* \mu^+ \mu^-$ at large recoil", *JHEP* **01** (2013) 048, arXiv:1207.2753.

[34] C. Bobeth *et al.*, "$B_{s,d} \to l^+ l^-$ in the Standard Model with Reduced Theoretical Uncertainty", *Phys. Rev. Lett.* **112** (2014) 101801, arXiv:1311.0903.

[35] LHCb and CMS collaborations, V. Khachatryan *et al.*, "Observation of the rare $B_s^0 \to \mu^+ \mu^-$ decay from the combined analysis of CMS and LHCb data", *Nature* **522** (2015) 68, arXiv:1411.4413.

[36] LHCb collaboration, R. Aaij *et al.*, "Measurement of the $B_s^0 \to \mu^+ \mu^-$ branching fraction and effective lifetime and search for $B^0 \to \mu^+ \mu^-$ decays", *Phys. Rev. Lett.* **118** (2017) 191801, arXiv:1703.05747.

[37] ATLAS collaboration, M. Aaboud *et al.*, "Study of the rare decays of B_s^0 and B^0 mesons into muon pairs using data collected during 2015 and 2016 with the ATLAS detector", *Submitted to: JHEP* (2018) arXiv:1812.03017.

[38] Belle collaboration, Y. Ushiroda *et al.*, "Time-Dependent CP Asymmetries in $B^0 \to K^0_S\pi^0\gamma$ transitions", *Phys. Rev.* **D74** (2006) 111104, arXiv:hep-ex/0608017.

[39] BaBar collaboration, B. Aubert *et al.*, "Measurement of Time-Dependent CP Asymmetry in $B^0 \to K^0_S\pi^0\gamma$ Decays", *Phys. Rev.* **D78** (2008) 071102, arXiv:0807.3103.

[40] Belle collaboration, J. Li *et al.*, "Time-dependent CP Asymmetries in $B^0 \to K^0_S\rho^0\gamma$ Decays", *Phys. Rev. Lett.* **101** (2008) 251601, arXiv:0806.1980.

[41] BaBar collaboration, P. del Amo Sanchez *et al.*, "Time-dependent analysis of $B^0 \to K^0_S\pi^-\pi^+\gamma$ decays and studies of the $K^+\pi^-\pi^+$ system in $B^+ \to K^+\pi^-\pi^+\gamma$ decays", *Phys. Rev.* **D93** (2016) 052013, arXiv:1512.03579.

[42] LHCb collaboration, R. Aaij *et al.*, "First experimental study of photon polarization in radiative B^0_s decays", *Phys. Rev. Lett.* **118** (2017) 021801, arXiv:1609.02032, [Addendum: Phys. Rev. Lett.118,no.10,109901(2017)].

[43] LHCb collaboration, R. Aaij *et al.*, "Observation of Photon Polarization in the $b \to s\gamma$ Transition", *Phys. Rev. Lett.* **112** (2014) 161801, arXiv:1402.6852.

[44] LHCb collaboration, R. Aaij *et al.*, "Angular analysis of the $B^0 \to K^{*0}e^+e$ decay in the low-q^2 region", *JHEP* **04** (2015) 064, arXiv:1501.03038.

[45] LHCb collaboration, R. Aaij *et al.*, "Differential branching fractions and isospin asymmetries of $B \to K^{(*)}\mu^+\mu^-$ decays", *JHEP* **06** (2014) 133, arXiv:1403.8044.

[46] LHCb collaboration, R. Aaij *et al.*, "Measurements of the S-wave fraction in $B^0 \to K^+\pi^-\mu^+\mu^-$ decays and the $B^0 \to K^*(892)^0\mu^+\mu^-$ differential branching fraction", *JHEP* **11** (2016) 047, arXiv:1606.04731, [Erratum: JHEP 04, (2017) 142].

[47] LHCb collaboration, R. Aaij *et al.*, "Angular analysis and differential branching fraction of the decay $B^0_s \to \phi\mu^+\mu^-$", *JHEP* **09** (2015) 179, arXiv:1506.08777.

[48] C. Bobeth, G. Hiller, D. van Dyk, and C. Wacker, "The Decay $B \to K\ell^+\ell^-$ at Low Hadronic Recoil and Model-Independent $\Delta B = 1$ Constraints", *JHEP* **01** (2012) 107, arXiv:1111.2558.

[49] R. R. Horgan, Z. Liu, S. Meinel, and M. Wingate, "Calculation of $B^0 \to K^{*0}\mu^+\mu^-$ and $B_s^0 \to \phi\mu^+\mu^-$ observables using form factors from lattice QCD", *Phys. Rev. Lett.* **112** (2014) 212003, arXiv:1310.3887.

[50] HPQCD, C. Bouchard *et al.*, "Standard Model Predictions for $B \to K\ell^+\ell^-$ with Form Factors from Lattice QCD", *Phys. Rev. Lett.* **111** (2013) 162002, arXiv:1306.0434, [Erratum: Phys. Rev. Lett.112,no.14,149902(2014)].

[51] A. Bharucha, D. M. Straub, and R. Zwicky, "$B \to V\ell^+\ell^-$ in the Standard Model from light-cone sum rules", *JHEP* **08** (2016) 098, arXiv:1503.05534.

[52] W. Altmannshofer and D. M. Straub, "New physics in $b \to s$ transitions after LHC run 1", *Eur. Phys. J.* **C75** (2015) 382, arXiv:1411.3161.

[53] LHCb collaboration, R. Aaij *et al.*, "Angular analysis of the $B^0 \to K^{*0}\mu^+\mu^-$ decay using 3 fb^{-1} of integrated luminosity", *JHEP* **02** (2016) 104, arXiv:1512.04442.

[54] ATLAS collaboration, M. Aaboud *et al.*, "Angular analysis of $B_d^0 \to K^*\mu^+\mu^-$ decays in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector", *JHEP* **10** (2018) 047, arXiv:1805.04000.

[55] Belle collaboration, S. Wehle *et al.*, "Lepton-Flavor-Dependent Angular Analysis of $B \to K^*\ell^+\ell^-$", *Phys. Rev. Lett.* **118** (2017) 111801, arXiv:1612.05014.

[56] CMS collaboration, A. M. Sirunyan *et al.*, "Measurement of angular parameters from the decay $B^0 \to K^{*0}\mu^+\mu^-$ in proton-proton collisions at $\sqrt{s} = 8$ TeV", *Phys. Lett.* **B781** (2018) 517, arXiv:1710.02846.

[57] S. Descotes-Genon, L. Hofer, J. Matias, and J. Virto, "On the impact of power corrections in the prediction of $B \to K^*\mu^+\mu^-$ observables", *JHEP* **12** (2014) 125, arXiv:1407.8526.

[58] LHCb collaboration, R. Aaij *et al.*, "Test of lepton universality using $B^+ \to K^+\ell^+\ell^-$ decays", *Phys. Rev. Lett.* **113** (2014) 151601, arXiv:1406.6482.

[59] LHCb collaboration, R. Aaij *et al.*, "Test of lepton universality with $B^0 \to K^{*0}\ell^+\ell^-$ decays", *JHEP* **08** (2017) 055, arXiv:1705.05802.

[60] BaBar collaboration, J. P. Lees *et al.*, "Measurement of Branching Fractions and Rate Asymmetries in the Rare Decays $B \to K^{(*)}l^+l^-$", *Phys. Rev.* **D86** (2012) 032012, arXiv:1204.3933.

[61] Belle collaboration, J. T. Wei *et al.*, "Measurement of the Differential Branching Fraction and Forward-Backword Asymmetry for $B \to K^{(*)}\ell^+\ell^-$", *Phys. Rev. Lett.* **103** (2009) 171801, arXiv:0904.0770.

[62] BaBar collaboration, J. P. Lees *et al.*, "Evidence for an excess of $\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau$ decays", *Phys. Rev. Lett.* **109** (2012) 101802, arXiv:1205.5442.

[63] BaBar collaboration, J. P. Lees *et al.*, "Measurement of an Excess of $\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau$ Decays and Implications for Charged Higgs Bosons", *Phys. Rev.* **D88** (2013) 072012, arXiv:1303.0571.

[64] Belle collaboration, M. Huschle *et al.*, "Measurement of the branching ratio of $\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau$ relative to $\bar{B} \to D^{(*)}\ell^-\bar{\nu}_\ell$ decays with hadronic tagging at Belle", *Phys. Rev.* **D92** (2015) 072014, arXiv:1507.03233.

[65] Belle collaboration, Y. Sato *et al.*, "Measurement of the branching ratio of $\bar{B}^0 \to D^{*+}\tau^-\bar{\nu}_\tau$ relative to $\bar{B}^0 \to D^{*+}\ell^-\bar{\nu}_\ell$ decays with a semileptonic tagging method", *Phys. Rev.* **D94** (2016) 072007, arXiv:1607.07923.

[66] Belle collaboration, S. Hirose *et al.*, "Measurement of the τ lepton polarization and $R(D^*)$ in the decay $\bar{B} \to D^*\tau^-\bar{\nu}_\tau$", *Phys. Rev. Lett.* **118** (2017) 211801, arXiv:1612.00529.

[67] LHCb collaboration, R. Aaij *et al.*, "Measurement of the ratio of branching fractions $\mathcal{B}(\bar{B}^0 \to D^{*+}\tau^-\bar{\nu}_\tau)/\mathcal{B}(\bar{B}^0 \to D^{*+}\mu^-\bar{\nu}_\mu)$", *Phys. Rev. Lett.* **115** (2015) 111803, arXiv:1506.08614, [Erratum: Phys. Rev. Lett.115,no.15,159901(2015)].

[68] LHCb collaboration, R. Aaij *et al.*, "Measurement of the ratio of the $B^0 \to D^{*-}\tau^+\nu_\tau$ and $B^0 \to D^{*-}\mu^+\nu_\mu$ branching fractions using three-prong τ-lepton decays", *Phys. Rev. Lett.* **120** (2018) 171802, arXiv:1708.08856.

[69] Heavy Flavor Averaging Group, Y. Amhis *et al.*, "Averages of b-hadron, c-hadron, and τ-lepton properties as of summer 2016", arXiv:1612.07233, updated results and plots available at http://www.slac.stanford.edu/xorg/hflav/.

[70] J. Lyon and R. Zwicky, "Resonances gone topsy turvy - the charm of QCD or new physics in $b \to s\ell^+\ell^-$?", arXiv:1406.0566.

[71] M. Ciuchini *et al.*, "$B \to K^*\ell^+\ell^-$ decays at large recoil in the Standard Model: a theoretical reappraisal", *JHEP* **06** (2016) 116, arXiv:1512.07157.

[72] V. G. Chobanova *et al.*, "Large hadronic power corrections or new physics in the rare decay $B \to K\mu^+\mu$?", *JHEP* **07** (2017) 025, arXiv:1702.02234.

[73] LHCb collaboration, R. Aaij *et al.*, "Measurement of the phase difference between short- and long-distance amplitudes in the $B^+ \to K^+\mu^+\mu^-$ decay", *Eur. Phys. J.* **C77** (2017) 161, arXiv:1612.06764.

[74] T. Blake *et al.*, "An empirical model to determine the hadronic resonance contributions $\overline{B}^0 \to \overline{K}^{*0}\mu^+\mu^-$ to transitions", *Eur. Phys. J.* **C78** (2018) 453, arXiv:1709.03921.

[75] M. Chrzaszcz *et al.*, "Prospects for disentangling long- and short-distance effects in the decays $B \to K^*\mu^+\mu^-$", arXiv:1805.06378.

[76] W. Altmannshofer, C. Niehoff, P. Stangl, and D. M. Straub, "Status of the $B \to K^*\mu^+\mu^-$ anomaly after Moriond 2017", *Eur. Phys. J.* **C77** (2017) 377, arXiv:1703.09189.

[77] W. Altmannshofer, P. Stangl, and D. M. Straub, "Interpreting Hints for Lepton Flavor Universality Violation", *Phys. Rev.* **D96** (2017) 055008, arXiv:1704.05435.

[78] B. Capdevila *et al.*, "Patterns of New Physics in $b \to s\ell^+\ell^-$ transitions in the light of recent data", *JHEP* **01** (2018) 093, arXiv:1704.05340.

[79] L.-S. Geng *et al.*, "Towards the discovery of new physics with lepton-universality ratios of $b \to s\ell\ell$ decays", *Phys. Rev.* **D96** (2017) 093006, arXiv:1704.05446.

[80] M. Ciuchini *et al.*, "On Flavourful Easter eggs for New Physics hunger and Lepton Flavour Universality violation", *Eur. Phys. J.* **C77** (2017) 688, arXiv:1704.05447.

[81] D. Du *et al.*, "Phenomenology of semileptonic B-meson decays with form factors from lattice QCD", *Phys. Rev.* **D93** (2016) 034005, arXiv:1510.02349.

[82] L. Evans and P. Bryant, "LHC Machine", *JINST* **3** (2008) S08001.

[83] LHCb collaboration, A. A. Alves Jr. *et al.*, "The LHCb detector at the LHC", *JINST* **3** (2008) S08005.

[84] W. Herr and B. Muratori, "Concept of luminosity", CERN-2006-002.

[85] CERN, "The CERN accelerator complex", http://cds.cern.ch/record/2636343, Accessed: 2018-09-16.

[86] W. Beenakker *et al.*, "QCD corrections to heavy quark production in hadron hadron collisions", *Nucl. Phys.* **B351** (1991) 507.

[87] E. Norrbin and T. Sjostrand, "Production and hadronization of heavy quarks", *Eur. Phys. J.* **C17** (2000) 137, arXiv:hep-ph/0005110.

[88] LHCb collaboration, R. Aaij *et al.*, "Measurement of the b-quark production cross-section in 7 and 13 TeV pp collisions", *Phys. Rev. Lett.* **118** (2017) 052002, arXiv:1612.05140, [Erratum: *Phys. Rev. Lett.* **119** (2017) 169901].

[89] LHCb collaboration, "$b\bar{b}$ production angle plots", https://lhcb.web.cern.ch/lhcb/speakersbureau/html/bb_ProductionAngles.html, Accessed: 2018-09-25.

[90] R. Aaij *et al.*, "Performance of the LHCb Vertex Locator", *JINST* **9** (2014) P09007, arXiv:1405.7808.

[91] R. Arink *et al.*, "Performance of the LHCb Outer Tracker", *JINST* **9** (2014) P01002, arXiv:1311.3893.

[92] LHCb Outer Tracker Group, P. d'Argent *et al.*, "Improved performance of the LHCb Outer Tracker in LHC Run 2", *JINST* **12** (2017) P11016, arXiv:1708.00819.

[93] LHCb collaboration, R. Aaij *et al.*, "LHCb detector performance", *Int. J. Mod. Phys.* **A30** (2015) 1530022, arXiv:1412.6352.

[94] LHCb collaboration, R. Aaij *et al.*, "Measurement of b-hadron masses", *Phys. Lett.* **B708** (2012) 241, arXiv:1112.4896.

[95] LHCb collaboration, R. Aaij *et al.*, "Measurement of the track reconstruction efficiency at LHCb", *JINST* **10** (2015) P02007, arXiv:1408.1251.

[96] LHCb RICH group, A. Papanestis and C. D'Ambrosio, "Performance of the LHCb RICH detectors during the LHC Run II", *Nucl. Instrum. Meth.* **A876** (2017) 221, arXiv:1703.08152.

[97] M. Adinolfi *et al.*, "Performance of the LHCb RICH detector at the LHC", *Eur. Phys. J.* **C73** (2013) 2431, arXiv:1211.6759.

[98] A. A. Alves Jr. *et al.*, "Performance of the LHCb muon system", *JINST* **8** (2013) P02022, arXiv:1211.1346.

[99] LHCb collaboration, B. Sciascia, "LHCb Run 2 trigger performance", *PoS* **BEAUTY2016** (2016) 029.

[100] LHCb collaboration, "Letter of Intent for the LHCb Upgrade", CERN-LHCC-2011-001.

[101] LHCb collaboration, "Framework TDR for the LHCb Upgrade: Technical Design Report", CERN-LHCC-2012-007.

[102] LHCb collaboration, "LHCb Tracker Upgrade Technical Design Report", CERN-LHCC-2014-001.

[103] LHCb collaboration, "LHCb VELO Upgrade Technical Design Report", CERN-LHCC-2013-021.

[104] LHCb collaboration, "LHCb PID Upgrade Technical Design Report", CERN-LHCC-2013-022.

[105] LHCb collaboration, "LHCb Trigger and Online Upgrade Technical Design Report", CERN-LHCC-2014-016.

[106] LHCb collaboration, "LHCb Upgrade Software and Computing Technical Design Report", CERN-LHCC-2018-007.

[107] LHCb collaboration, "Expression of Interest for a Phase-II LHCb Upgrade: Opportunities in flavour physics, and beyond, in the HL-LHC era", CERN-LHCC-2017-003.

[108] LHCb collaboration, I. Bediaga *et al.*, "Physics case for an LHCb Upgrade II - Opportunities in flavour physics, and beyond, in the HL-LHC era", arXiv:1808.08865.

[109] LHCb collaboration, R. Aaij *et al.*, "First measurement of the differential branching fraction and *CP* asymmetry of the $B^\pm \to \pi^\pm \mu^+ \mu^-$ decay", *JHEP* **10** (2015) 034, arXiv:1509.00414.

[110] LHCb collaboration, R. Aaij *et al.*, "Observation of the suppressed decay $\Lambda_b^0 \to p\pi^- \mu^+ \mu^-$", *JHEP* **04** (2017) 029, arXiv:1701.08705.

[111] LHCb collaboration, R. Aaij *et al.*, "Study of the rare B_s^0 and B^0 decays into the $\pi^+\pi^-\mu^+\mu^-$ final state", *Phys. Lett.* **B743** (2015) 46, arXiv:1412.6433.

[112] Y.-L. Wu, M. Zhong, and Y.-B. Zuo, "$B_{(s)}$, $D_{(s)} \to \pi$, K, η, ρ, K^*, ω, ϕ transition form factors and decay rates with extraction of the CKM parameters $|V_{ub}|$, $|V_{cs}|$, $|V_{cd}|$", *Int. J. Mod. Phys.* **A21** (2006) 6125, arXiv:hep-ph/0604007.

[113] R. N. Faustov and V. O. Galkin, "Rare B_s decays in the relativistic quark model", *Eur. Phys. J.* **C73** (2013) 2593, arXiv:1309.2160.

[114] B. Kindra and N. Mahajan, "Predictions of angular observables for $\bar{B}_s \to K^* \ell\ell$ and $\bar{B} \to \rho\ell\ell$ in Standard Model", arXiv:1803.05876.

[115] S. S. Wilks, "The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses", *Annals Math. Statist.* **9** (1938) 60.

[116] Belle collaboration, K. Chilikin *et al.*, "Observation of a new charged charmonium like state in $\bar{B}^0 J/K^{-+}$ decays", *Phys. Rev.* **D90** (2014) 112009, arXiv:1408.6457.

[117] LHCb collaboration, "Updated average f_s/f_d b-hadron production fraction ratio for 7 TeV pp collisions", LHCb-CONF-2013-011.

[118] W. D. Hulsbergen, "Decay chain fitting with a Kalman filter", *Nucl. Instrum. Meth.* **A552** (2005) 566, arXiv:physics/0503191.

[119] M. Clemencic *et al.*, "The LHCb simulation application, Gauss: Design, evolution and experience", *J. Phys. Conf. Ser.* **331** (2011) 032023.

[120] T. Sjöstrand, S. Mrenna, and P. Skands, "A brief introduction to PYTHIA 8.1", *Comput. Phys. Commun.* **178** (2008) 852, arXiv:0710.3820.

[121] I. Belyaev *et al.*, "Handling of the generation of primary events in Gauss, the LHCb simulation framework", *J. Phys. Conf. Ser.* **331** (2011) 032047.

[122] D. J. Lange, "The EvtGen particle decay simulation package", *Nucl. Instrum. Meth.* **A462** (2001) 152.

[123] P. Golonka and Z. Was, "PHOTOS Monte Carlo: A precision tool for QED corrections in Z and W decays", *Eur. Phys. J.* **C45** (2006) 97, arXiv:hep-ph/0506026.

[124] Geant4 collaboration, S. Agostinelli *et al.*, "Geant4: A simulation toolkit", *Nucl. Instrum. Meth.* **A506** (2003) 250; Geant4 collaboration, J. Allison *et al.*, "Geant4 developments and applications", *IEEE Trans.Nucl.Sci.* **53** (2006) 270.

[125] M. Pivk and F. R. Le Diberder, "SPlot: A Statistical tool to unfold data distributions", *Nucl. Instrum. Meth.* **A555** (2005) 356, arXiv:physics/0402083.

[126] LHCb collaboration, R. Aaij *et al.*, "Measurement of the fragmentation fraction ratio f_s/f_d and its dependence on B meson kinematics", *JHEP* **04** (2013) 001, arXiv:1301.5286.

[127] M. Feindt and U. Kerzel, "The NeuroBayes neural network package", *Nucl. Instrum. Meth.* **A559** (2006) 190.

[128] L. Anderlini *et al.*, "The PIDCalib package", CERN-LHCB-PUB-2016-021.

[129] LHCb collaboration, R. Aaij *et al.*, "Measurement of CP asymmetries and polarisation fractions in $B^0_s \to K^{*0}\bar{K}^{*0}$ decays", *JHEP* **07** (2015) 166, arXiv:1503.05362.

[130] T. Skwarnicki, "A study of the radiative CASCADE transitions between the Upsilon-Prime and Upsilon resonances", PhD thesis, Institute of Nuclear Physics, Cracow, 1986.

[131] LHCb collaboration, R. Aaij *et al.*, "Measurement of the polarization amplitudes in $B^0 \to J/\psi K^*(892)^0$ decays", *Phys. Rev.* **D88** (2013) 052002, arXiv:1307.2782.

[132] LHCb collaboration, R. Aaij *et al.*, "Measurement of CP violation parameters and polarisation fractions in $B^0_s \to J/\psi \overline{K}^{*0}$ decays", *JHEP* **11** (2015) 082, arXiv:1509.00400.

[133] LHCb collaboration, R. Aaij *et al.*, "Measurement of the $B^0_s \to J/\psi \overline{K}*0$ branching fraction and angular amplitudes", *Phys. Rev.* **D86** (2012) 071102, arXiv:1208.0738.

[134] A. C. Crocombe, T. E. Latham, A. Mathad, and M. R. J. Williams, "Geometry descriptions and performance studies of the VELO upgrade module support candidate designs", Tech. Rep. LHCb-INT-2018-025, CERN, Geneva, Sep, 2018.

[135] A. C. Crocombe, T. E. Latham, and M. R. J. Williams, "Geometry descriptions and physics performance of the VELO Upgrade module cooling candidate designs", Tech. Rep. LHCb-INT-2018-026, CERN, Geneva, Sep, 2018.